BANK OF CANADA
BANQUE DU CANADA

Working Paper/Document de travail
2015-17

Testing for the Diffusion Matrix in a Continuous-Time Markov Process Model with Applications to the Term Structure of Interest Rates

by Fuchun Li

www.bank-banque-canada.ca

Bank of Canada Working Paper 2015-17

June 2015

Testing for the Diffusion Matrix in a Continuous-Time Markov Process Model with Applications to the Term Structure of Interest Rates

by

Fuchun Li

Financial Stability Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
fuchunli@bankofcanada.ca

Bank of Canada working papers are theoretical or empirical works-in-progress on subjects in economics and finance. The views expressed in this paper are those of the author. No responsibility for them should be attributed to the Bank of Canada.

ISSN 1701-9397

© 2015 Bank of Canada

Acknowledgements

I would like to thank seminar participants at the Bank of Canada, Carleton University and Monash University, as well as Federico Bandi, James Chapman, Bruno Feunou, JianJian Jin, Dennis Kristensen and Yasuo Terajima for their comments and suggestions. I also thank Yacine Aït-Sahalia and Greg Duffee for providing me with the eurodollar interest rates and U.S. Treasury yield data, respectively, as well as Hong Xiao for excellent research assistance.

Abstract

The author proposes a test for the parametric specification of each component in the diffusion matrix of a d-dimensional diffusion process. Overall, $d(d\text{-}1)/2$ test statistics are constructed for the off-diagonal components, while d test statistics are constructed for the main diagonal components. Using theories of degenerate U-statistics, each of these test statistics is shown to follow an asymptotic standard normal distribution under null hypothesis, while diverging to infinity if the component is misspecified over a significant range. Our tests strongly reject the specification of diffusion functions in a variety of popular univariate interest rate models for daily 7-day eurodollar spot rates, and the specification of the diffusion matrix in some popular multivariate affine term-structure models for monthly U.S. Treasury yields.

JEL classification: C12, C14, E17, E43, G12, G20
Bank classification: Asset pricing; Interest rates; Econometric and statistical methods

Résumé

L'auteur propose un test permettant de vérifier la validité de la spécification paramétrique des différentes composantes de la matrice de distribution d'un processus de distribution à d dimensions. À cette fin, il construit $d(d\text{-}1)/2$ statistiques de test pour les composantes hors-diagonale, et d statistiques pour les composantes de la diagonale principale. En se fondant sur les théories des U-statistiques dégénérées, l'auteur montre que chacune de ces statistiques de test suit asymptotiquement une loi de distribution normale sous l'hypothèse nulle, mais diverge à l'infini si la spécification de la composante est erronée sur une large fourchette. Ses tests invalident clairement la spécification des fonctions de distribution de divers modèles univariés de taux d'intérêt, très utilisés, lorsque ces modèles sont appliqués aux taux au comptant pour les dépôts à sept jours en eurodollars, ainsi que la spécification de la matrice de distribution utilisée dans certains modèles affines multivariés de la structure par terme, également très utilisés pour les rendements mensuels des titres du Trésor américain.

Classification JEL : C12, C14, E17, E43, G12, G20
Classification de la Banque : Évaluation des actifs; Taux d'intérêt; Méthodes économétriques et statistiques

Non-technical Summary

Continuous-time Markov process models are powerful analytic tools in economics and finance for studying issues such as asset pricing, the decision to optimally consume, portfolio choice under a variety of constraints, game theory and contract theory, etc. As with many model specifications in economics and finance, since economic theory does not suggest a particular functional form for a continuous-time model, the specification of a continuous-time model is usually based on analysis, convenience and the empirical experience of the practitioner. As a consequence, a possibly serious problem with the specification of a continuous-time model is model misspecification, which could lead to misleading results flowing from its implications for financial analysis and statistical inference. A test is thus required to determine whether a continuous-time model can appropriately capture the dynamics implied by the data.

At the same time, continuous-time models in economics and finance, which until recently have been largely univariate, now predominantly include multiple state variables. Typical examples include asset-pricing models with multiple explanatory factors, term-structure models with multiple yields or factors, and stochastic volatility or stochastic mean reversion models. Motivated by this trend and the fact that the diffusion matrix in such a model is a crucial factor in modeling the movements of individual state variables, such as interest rates, asset prices or exchange rates, and the comovements among state variables, we propose a test for the parametric specification of each component in the diffusion matrix. Monte Carlo simulations show that our tests have satisfactory size and power performance.

To highlight our tests, we apply our tests to popular multivariate affine term-structure models, obtaining many interesting new empirical findings.

1 Introduction

Continuous-time Markov process models (diffusion process models) are powerful analytic tools in economics and finance to study issues such as asset pricing, the decision to optimally consume, portfolio choice under a variety of constraints, such as game theory, contract theory, etc. As with many model specifications in economics and finance, since economic theory does not suggest a particular functional form for a diffusion process, the specification of a diffusion process is usually based on analysis, convenience and the empirical experience of the practitioner. As a consequence, a possibly serious problem with the specification of a diffusion process is model misspecification, which could lead to misleading results from its implications for financial analysis and statistical inference. A test is thus required to determine whether a diffusion process can appropriately capture the dynamics implied by the data.

Much progress has been made in testing a parametric specification of a diffusion process in recent years. For instance, Chen and Hong (2010); Aït-Sahalia, Fan, and Peng (2009); Thompson (2008); Chen, Gao, and Tang (2008); Corradi and Swanson (2005); Hong and Li (2005); and Diebold, Gunther, and Tay (1998) proposed tests based on a comparison of a nonparametric estimation of the conditional characteristic function, the density function, the transition density or the conditional distribution function with their corresponding parametric counterparts assumed by the null hypothesis. A significant limitation of these tests is that when a misspecified diffusion process is rejected, they cannot identify whether the model misspecification comes from the drift vector or the diffusion matrix.[1] However, in order to check possible sources of model misspecification, such information is of crucial importance in reconstructing a diffusion model.

[1] The properties of a diffusion process are determined entirely by the drift vector and the diffusion matrix. Thus, the problem of selecting among competing diffusion process models or specifying an alternative diffusion process model comes down to choosing the drift vector and diffusion matrix.

Li (2007) proposed a test for the parametric specification of the diffusion function in a univariate diffusion process.[2] The most important feature of this test is that it can directly detect whether the diffusion function is correctly specified even if there is no information about the functional form of the drift function. The limitation of this test is that it applies only to a univariate diffusion process. In many cases, a multivariate diffusion process is needed to study the dynamic behavior of multiple state variables. For instance, multivariate continuous-time models for the dynamic term structures of interest rates (Dai and Singleton, 2000; Piazzesi, 2010) and equity returns (Andersen, Benzoni and Lund, 2002) have been developed and widely used in pricing and hedging fixed-income or equity derivatives, managing financial risk, and evaluating monetary policy and debt policy. Specifically, the diffusion matrix in such a model, as the second moment and the measurement of the instantaneous volatility of the state variables, is a crucial factor in modeling the movements of individual state variables, such as interest rates, asset prices or exchange rates, and the comovements among state variables. [3] An extension of the test in Li (2007) to multivariate diffusion processes is thus needed to detect the specification of the diffusion matrix because each component in the diffusion matrix differs fundamentally in its implications for capturing the stochastic behavior of the state variables.

The objective of this paper is to test the validity of the parametric specification of the diffusion matrix in a multivariate diffusion process without any restrictions on the functional form of the drift vector. For this reason, we confirm whether a diffusion matrix is correctly specified by testing whether each component in this diffusion matrix is correctly specified. [4] For each

[2] In an univariate diffusion process, the drift vector and diffusion matrix are called the drift function and the diffusion function, respectively.

[3] A number of theoretical studies and empirical evidence (Aït-Sahalia, 1996; Durham, 2003) have concluded that the diffusion matrix plays a very important role in predicting the movements of derivative security prices, determining optimal portfolio hedging strategies for risk-averse investors, or creating some leverage within a portfolio. A more intuitive example is that, in the famous Black-Scholes option-pricing formula, the prices of derivative securities are affected by the price of underlying assets only through its diffusion function.

[4] Another potential approach to testing the parametric specification of the diffusion matrix would be to follow the

component in the diffusion matrix of a d-dimensional diffusion process, we propose a test of the parametric specification of this component based on the comparison between the model-implied parametric specification of the component and a nonparametric estimator of the component. Since the diffusion matrix has $d(d+1)/2$ different components, $d(d+1)/2$ different test statistics are constructed. Consequently, not only can these tests detect whether the diffusion matrix is specified correctly, but they can also provide a clear indication of the direction of the misspecification in the diffusion matrix; that is, which components in this matrix are possibly misspecified and which are not. It turns out that when the component is on the main diagonal, the test statistic is a natural extension of the test statistic in Li (2007) even if there exist correlation effects between state variables; when the component is off the main diagonal, the test statistic is characterized in a different way.

Using the seven-day eurodollar interest rate data, as in Aït-Sahalia (1996), we apply the test in Li (2007) to evaluate the specification of the diffusion functions of five popular univariate interest rate models.[5] Aït-Sahalia's test (1996) rejects all linear drift models, but it would not reject Aït-Sahalia's nonlinear drift model (1996). In contrast, our test firmly rejects the specification of the diffusion function in each of these univariate diffusion models. The diffusion function in the Chan et al. (1992) model has the best performance. Using monthly U.S. Treasury yields, as in Duffee (2002) and Hong and Li (2005), our tests reject the affine specification of the diffusion matrix in the three-factor diffusion process of yields, which implies the joint misspecification of the drift vector and diffusion matrix in the affine term-structure model. Meanwhile, the affine model characterized

methodology that White (1980) and Eklund and Teräsvirta (2007) used to test the constancy of the error covariance matrix in a regression model by comparing the estimator of a model-implied diffusion matrix with a nonparametric estimator of the diffusion matrix.

[5]The Aït-Sahalia (1996) test is based on the comparison between the marginal density implied by a parametric diffusion process and the density estimated nonparametrically. Although the seven-day eurodollar interest rate data are used for extensive analysis in many papers (Durham, 2003), empirically testing for the specification of the diffusion functions in the models for the term structure of interest rates remains an unanswered question.

by the assumption that the volatilities of state variables are determined by two of the three state variables performs the best.

This paper is organized as follows. Section 2 states the model, hypotheses of interest, and our test statistic for each component in the diffusion matrix. In Section 3, we derive the asymptotic null distribution of each of these test statistics, and discuss its asymptotic power property. In Section 4, we assess the finite sample performance of our test statistics using Monte Carlo studies. In Section 5, we evaluate the performance of the diffusion functions in five popular univariate interest rate models for seven-day eurodollar interest rates, and the performance of the diffusion matrix in some popular multivariate affine term-structure models for monthly U.S. Treasury yields. Section 6 concludes. All proofs are provided in the Appendix.

2 Model, Hypotheses and Test Statistics

Suppose that a state vector x_t follows the dynamics,

$$x_t = x_{t_0} + \int_{t_0}^{t} \mu(x_s)ds + \int_{t_0}^{t} \sigma(x_s)dB_s, t_0 \leq t < \infty, \tag{1}$$

where x_{t_0} is a given initial condition, $x_t \equiv (x_t^1, ..., x_t^d)'$ and $\mu(x_t) \equiv (\mu_1(x_t), ..., \mu_d(x_t))'$ are $d \times 1$ vectors, $\sigma(x_t) \equiv \{\sigma_{ij}(x_t)\}_{1 \leq i,j \leq d}$ is a $d \times d$ matrix, and $B_t \equiv (B_t^1, ..., B_t^d)'$ is a d-dimensional vector of independent standard Brownian motions.[6] Assume that x_{t_0} is independent of B_t.

For $1 \leq i \leq d$, the coordinate x_t^i of the stochastic differential equation (1) can be written as,

$$x_t^i = x_{t_0}^i + \int_{t_0}^{t} \mu_i(x_s)ds + \sum_{j=1}^{d} \int_{t_0}^{t} \sigma_{ij}(x_s)dB_s^j, t_0 \leq t < \infty. \tag{2}$$

Let $x \equiv (x^1, ..., x^d) \in R^d$, then we define the $d \times d$ symmetric and non-negative diffusion matrix $a(x) \equiv \sigma(x)\sigma(x)'$ with the general element $a_{ij}(x) = \sum_{l=1}^{d} \sigma_{il}(x)\sigma_{jl}(x), 1 \leq i, j \leq d$.

[6] Independence is just without loss of generality because any correlation structures between the shocks to the different equations can be modeled through the inclusion of off-diagonal terms in the $\sigma(\cdot)$ matrix.

The dynamic properties of process x_t are characterized by its transition density function, which depends on $\mu(\cdot)$ and $a(x)$.[7] In fact, it can be shown that, if there exists a continuum of solutions in $\sigma(\cdot)$ to equation $a(x) \equiv \sigma(x)\sigma(x)'$, then the transition probability function is identical for each of these $\sigma(x)$ (Aït-Sahalia, 2008; and Remark 5.1.7 and Section 5.3 in Stroock and Varadhan, 1979).

We use $\{a^0(x,\theta) : \theta \in \Theta\}$, where Θ is a subset of R^p, to denote the model-implied parametric family of diffusion matrix $a^0(x,\theta) = \sigma^0(x,\theta)\sigma^0(x,\theta)'$ with general element $a^0_{i,j}(x,\theta) = \sum_{l=1}^{d} \sigma^0_{il}(x,\theta)\sigma^0_{jl}(x,\theta), 1 \le i,j \le d$. Supposing that we do not have any information on the functional form of the drift vector, we want to test whether the true diffusion matrix $a(x)$ belongs to the given parametric family $\{a^0(x,\theta) : \theta \in \Theta\}$.

It is obvious that $a(x) = a^0(x,\theta_0)$ almost everywhere for some $\theta_0 \in \Theta$ if and only if there exists a parameter $\theta_0 \in \Theta$ such that $a_{ij}(x) = a^0_{ij}(x,\theta_0)$ almost everywhere for any $1 \le i \le j \le d$. Hence, our approach is to test the parametric specification of the diffusion matrix by testing the null hypothesis that for any $1 \le i \le j \le d$, there exists a parameter $\theta_0 \in \Theta$ such that $a_{ij}(x) = a^0_{ij}(x,\theta_0)$ almost everywhere, versus the alternative, that $a_{ij}(x) \ne a^0_{ij}(x,\theta)$ with a positive measure for any $\theta \in \Theta$, that is,

$$H_0^{ij} : a_{ij}(x) = a^0_{ij}(x,\theta_0), \text{almost everywhere for some } \theta_0 \in \Theta, \tag{3}$$

versus the alternative hypothesis,

$$H_A^{ij} : a_{ij}(x) \ne a^0_{ij}(x,\theta), \text{on a subset with a positive measure for any } \theta \in \Theta. \tag{4}$$

Corresponding to $a^0_{ij}(x,\theta_0)$, the parametric specification of the component $a_{ij}(x)$, our test is based on the integrated squared difference between $a_{ij}(x)$ and $a^0_{ij}(x,\theta_0)$,

$$I_{ij} \equiv E\{[(a_{ij}(x_t) - a^0_{ij}(x_t,\theta_0))f(x_t)]^2 w(x_t)\}, \tag{5}$$

[7] The Kolmogorov forward and backward equations characterize the transition density of a diffusion process, and are determined by the drift vector and diffusion matrix.

where $f(x)$ is the unknown marginal density of x_t, and is used to trim the small values of the random denominator in the nonparametric estimation of $a_{ij}(x)$, and $w(x)$ is the weighting function. In the literature on constructing a test for model specification in economics and finance, the distance measures similar to (5) are used to test for a parametric characteristic function, a parametric transition density function, a parametric regression function, and a parametric diffusion function by, for instance, Chen and Hong (2010), Aït-Sahalia, Fan and Peng (2009), Aït-Sahalia, Bickel and Stoker (2001), and Li (2007). The weighting function in (5) can be used to remove extreme observations or outliers to obtain asymptotic results, and can allow us to focus on a particular empirical question of interest and reduce the influences of unreliable estimates.

Note that $I_{ij} \geq 0$ and the equality holds if and only if H_0^{ij} is true. Hence, I_{ij} can serve as a proper candidate for consistently testing H_0^{ij}. Although we use the same measure for different components in the diffusion matrix, it will turn out that the test statistic for a different component in the diffusion matrix is characterized differently. To get a feasible test statistic, we need to estimate $a_{ij}(x), \theta_0$, and $f(x)$ in (5). Suppose that the process x_t is observed at equi-spaced times $\{t = t_1, t_2, ..., t_n\}$ in the time interval $[t_0, T]$, where T is a strict positive number. These observations can be expressed as $\{x_t = x_{t_0+\triangle_n}, x_{t_0+2\triangle_n}, ..., x_{t_0+n\triangle_n}\}$ at $\{t_1 = t_0 + \triangle_n, t_2 = t_0 + 2\triangle_n, ..., t_n = t_0 + n\triangle_n\}$, where $\triangle_n = (T - t_0)/n$ is the sampling interval. We use the notation $x_{n,t}$ to express the observation on the process x_t at $t_0 + t\triangle_n$, i.e., $x_{n,t} \equiv x_{t_0+t\triangle_n}$. Therefore, the data are given by a triangular array of random variables $\{x_{n,t}, 1 \leq t \leq n\}$, where $x_{n,t} \equiv (x_{n,t}^1, ..., x_{n,t}^d)$.[8]

Under both H_0^{ij} and H_A^{ij}, the true and unknown $a_{ij}(x)$ can be estimated by the kernel method,

[8]Since the data are given by a triangular array of random variables, after establishing the test statistics, we will need to use a central limit theorem for degenerate U-statistics of the triangular array of random variables to show the asymptotic distributions of these test statistics.

which is proposed by Bandi and Moloche (2008),

$$\hat{a}_{ij}(x) \equiv \frac{\sum_{t=1}^{n-1} K(\frac{x_{n,t}-x}{h_n})[x^i_{n,t+1}-x^i_{n,t}][x^j_{n,t+1}-x^j_{n,t}]}{\triangle_n \sum_{t=1}^{n-1} K(\frac{x_{n,t}-x}{h_n})}, \tag{6}$$

where $K(\cdot)=\Pi_{i=1}^{d}k(\cdot)$ is a product kernel function and h_n is a smoothing parameter whose properties will be presented in the next section. θ_0 can be estimated by the minimum contrast estimator,

$$\hat{\theta}_n = argmin_{\theta\in\Theta}\frac{1}{n}\sum_{t=1}^{n-1}[\log\det(a^0(x_{n,t},\theta)) + \frac{1}{\triangle_n}(x_{n,t+1}-x_{n,t})(a^0(x_{n,t},\theta))^{-1}(x_{n,t+1}-x_{n,t})'], \tag{7}$$

which is proposed by Genon-Catalot and Jacod (1993), or the minimizing distance estimator,

$$\hat{\theta}_n = argmin_{\theta\in\Theta}\sum_{t=1}^{n}|\hat{a}_{ij}(x_{n,t}) - a^0(x_{n,t},\theta)|, \tag{8}$$

which is proposed by Bandi and Phillips (2007). Note that both estimation methods can be used for the estimation of the parameter θ_0 in the diffusion matrix without any information about the functional form of the drift vector. The density function $f(x)$ can be estimated by the kernel method,

$$\hat{f}(x) = \frac{1}{nh_n}\sum_{t=1}^{n-1}K(\frac{x_{n,t}-x}{h_n}). \tag{9}$$

Inserting (6), either (7) or (8), and (9) into the definition of I_{ij} given by (5), yields the following estimator of I_{ij},

$$I_{nij} = \frac{1}{n}\sum_{t=1}^{n}[(\hat{a}_{ij}(x_{n,t}) - a^0_{ij}(x_{n,t},\hat{\theta}_n))\hat{f}(x_{n,t})]^2 w(x_{n,t}). \tag{10}$$

The test statistic for H_0^{ij} versus H_A^{ij} is an appropriately centered and scaled version of I_{ij}, namely,

$$J_{nij} \equiv nh^{d/2}[I_{nij} - r_{nij}(d)]/v_{nij}, \tag{11}$$

where

$$v^2_{nij} = \frac{2}{n}\sum_{t=1}^{n}[\hat{a}_{ii}(x_{n,t})\hat{a}_{jj}(x_{n,t}) + \hat{a}^2_{ij}(x_{n,t})]^2\hat{f}^3(x_{n,t})w^2(x_{n,t})\int[\int K(u)K(v+u)du]^2dv, \tag{12}$$

and $r_{nij}(d)$ is the estimator of the asymptotic bias term, and can be expressed as,

if $d \leq 3$,

$$r_{nij}(d) = \frac{1}{n^2 h_n^d} \sum_{t=1}^{n} [\hat{a}_{ii}(x_{n,t})\hat{a}_{jj}(x_{n,t}) + \hat{a}_{ij}^2(x_{n,t})] w(x_{n,t}) \hat{f}(x_{n,t}) \int K^2(u) du, \tag{13}$$

if $d \geq 4$,

$$\begin{aligned}
r_{nij}(d) &= \frac{1}{n^2 h_n^d} \sum_{t=1}^{n} [\hat{a}_{ii}(x_{n,t})\hat{a}_{jj}(x_{n,t}) + \hat{a}_{ij}^2(x_{n,t})] w(x_{n,t}) \hat{f}(x_{n,t}) \int K^2(u) du \\
&+ \frac{1}{n^2 h_n^d} \sum_{l=2}^{m-1} \frac{h_n^l}{l!} \sum_{t=1}^{n} \int [(u'\nabla)^l (\hat{a}_{ii}(x_{n,t})\hat{a}_{jj}(x_{n,t}) \hat{f}(x_{n,t}))] w(x_{n,t}) K^2(u) du \\
&+ 2\frac{1}{n^2 h_n^d} \sum_{l=2}^{m-1} \frac{h_n^l}{l!} \sum_{t=1}^{n} \int [(u'\nabla)^l (\hat{a}_{ij}^2(x_{n,t}) \hat{f}(x_{n,t}))] w(x_{n,t}) K^2(u) du \\
&- 2\frac{1}{n^2 h_n^d} \sum_{l=2}^{m-1} \frac{h_n^l}{l!} \sum_{t=1}^{n} \int [(u'\nabla)^l (\hat{a}_{ij}(x_{n,t}) \hat{f}(x_{n,t}))] \hat{a}_{ij}(x_{n,t}) w(x_{n,t}) K^2(u) du \\
&+ \frac{1}{n^2 h_n^d} \sum_{l=2}^{m-1} \frac{h_n^l}{l!} \sum_{t=1}^{n} \int [(u'\nabla)^l \hat{f}(x_{n,t})] \hat{a}_{ij}^2(x_{n,t}) w(x_{n,t}) K^2(u) du,
\end{aligned} \tag{14}$$

where $m = [d/2] + 1$, $[x]$ is the greatest integer function of a real number x, $\nabla = (\frac{\partial}{\partial x^1}, ..., \frac{\partial}{\partial x^d})$, and $(u'\nabla)^l g(x) \equiv (u_1 \frac{\partial}{\partial x^1} + ... + u_d \frac{\partial}{\partial x^d})^l g(x)$ for $l = 1, 2, ..., m-1$ and any l times differentiable function $g(x)$. When $d \geq 4$, we need to estimate the derivatives of $\hat{a}_{ij}(x)$. From (6), we have,

$$\begin{aligned}
\frac{\partial \hat{a}_{ij}(x)}{\partial x^l} &= \frac{\hat{a}_{ij}(x) \frac{1}{nh_n^{d+1}} \sum_{t=1}^{n-1} K_{x^l}(\frac{x_{n,t}-x}{h_n})}{\frac{1}{nh_n^d} \sum_{t=1}^{n-1} K(\frac{x_{n,t}-x}{h_n})} \\
&\quad - \frac{\frac{1}{nh_n^{d+1} \triangle_n} \sum_{t=1}^{n-1} K_{x^l}(\frac{x_{n,t}-x}{h_n}) [x_{n,t+1}^i - x_{n,t}^i][x_{n,t+1}^j - x_{n,t}^j]}{\frac{1}{nh_n^d} \sum_{t=1}^{n-1} K(\frac{x_{n,t}-x}{h_n})},
\end{aligned} \tag{15}$$

where $1 \leq l \leq d$, and $K_{x^l}(\frac{x_{n,t}-x}{h_n}) = \frac{\partial K(\frac{x_{n,t}-x}{h_n})}{\partial x^l}$.

In general, for any positive integer $L \leq m-1$, we have,

$$\begin{aligned}
\frac{\partial^L \hat{a}_{ij}(x)}{\partial x^{d_1} ... \partial x^{d_L}} &= \frac{\sum_{l=1}^{L} \frac{\partial^{L-l}}{\partial x^{d_{l+1}} ... \partial x^{d_L}} [\frac{1}{nh_n^{d+1}} \sum_{t=1}^{n-1} K_{x^{d_l}}(\frac{x_{n,t}-x}{h_n}) \frac{\partial^{l-1} \hat{a}_{ij}(x)}{\partial x^{d_1} ... x^{d_{l-1}}}]}{\frac{1}{nh_n^d} \sum_{t=1}^{n-1} K(\frac{x_t-x}{h_n})} \\
&\quad + \frac{\frac{(-1)^L}{nh_n^{d+L} \triangle_n} \sum_{t=1}^{n-1} \frac{\partial^L K(\frac{x_t-x}{h_n})}{\partial x^{d_1} ... \partial x^{d_L}} (x_{n,t+1}^i - x_{n,t}^i)(x_{n,t+1}^j - x_{n,t}^j)}{\frac{1}{nh_n^d} \sum_{t=1}^{n-1} K(\frac{x_t-x}{h_n})},
\end{aligned} \tag{16}$$

where $1 \leq d_i \leq d$, for $i = 1, 2, ..., L$.

Note that $r_{nij}(d)$ in (14) is an estimator of the asymptotic bias term $\bar{I}_{nij}^{12}$ in (A.22) in the Appendix, and it is reduced to (13) when $d \leq 3$ by ignoring higher-order terms than $n^{-2}h_n^d$ in (14). This is because when $d \leq 3$, ignoring these higher-order terms does not impact the asymptotic distribution of our test statistic, J_{nij}, under the assumptions in Section 3. Since the diffusion matrix has $d(d+1)/2$ different components, $d(d+1)/2$ different test statistics are constructed. Any component in the drift vector enters these test statistics; therefore, these test statistics can be used to test the parametric specification of the diffusion matrix even if there is no information on the function form of the drift vector. Since the test statistic J_{nij} does not contain another component, with the exception of $a_{ij}^0(x, \hat{\theta}_0)$ in the diffusion matrix $a_0(x, \theta)$, J_{nij} can be used to test the null hypothesis H_0^{ij} even if there exist misspecifications for any other components in the diffusion matrix $a^0(x, \theta)$. Moreover, when the component, for instance, $a_{ii}(x)$, is on the main diagonal, the test statistic J_{nii} only depends on the nonparametric estimator of $a_{ii}(x)$ and the parametric estimator of $a_{ii}^0(x, \theta)$. Thus, it is a natural extension of the test statistic in Li (2007) even if correlation effects exist in the true diffusion matrix. When the component $a_{ij}^0(x, \theta)$ is off the main diagonal ($i \neq j$), the test statistic J_{nij} depends not only on the nonparametric estimator of $a_{ij}(x)$ and the parametric estimator of $a_{ij}^0(x, \theta)$, but also on the nonparametric estimators of $a_{ii}(x)$ and $a_{jj}(x)$. As a consequence, the test statistic J_{nij} is characterized in a different way than J_{nii}.

3 Asymptotic Distributions

For the purpose of asymptotic analysis, we make the following assumptions.

Assumption 1. *Let $D = \prod_{i=1}^{d}(l_i, r_i)$ be a product of d intervals, (l_i, r_i), for $i = 1, ..., d$, where $-\infty \leq l_i < r_i \leq \infty$. On D, the functions $\mu(x)$ and $\sigma(x)$ are continuously differentiable.*

Assumption 2. *$\mu(x)$ and $\sigma(x)$ satisfy the following conditions.*

(i) There exists a positive constant C_D such that for every $x \in D$,

$$|\mu(x)|^2 + |\sigma(x)|^2 \leq C_D(1+|x|^2), \tag{17}$$

where $|\cdot|$ denotes the Max norm for the matrix.

(ii) There exists a nonnegative function $\rho(\cdot,\cdot)$ such that $E[\rho^4(x_{t_1},x_{t_2})] \leq C$ for any $t_1,t_2 \in [t_0,T]$, where C is a positive constant, and

$$|\mu(x)-\mu(y)| \leq \rho(x,y)|x-y|. \tag{18}$$

(iii) $a(x)$ is positively definite for all x in the interior of D, and the partial derivatives of $a_{ij}(x)$ up to the order of r have the first absolute moment, where r is the order for the kernel function that will be defined in Assumption 5. There exists an $\bar{r}^{th}$-degree homogeneous polynomial $Q_{ij}(y,x)$ in $(y-x)$, such that $sup_{y\in A_x^b}|a_{ij}(y)-a_{ij}(x)-Q_{ij}(y,x)|/|y-x|^{\alpha} \leq g(x)$ for all $x \in D$, where $\bar{r} \leq r$ and $\bar{r} \leq \alpha \leq \bar{r}+1, A_x^b = \{y : |y-x| < b\}, b > 0$. The coefficients of $Q_{ij}(y,x)$ are the partial derivatives of $a_{ij}(x)$ at x up to the order of $\bar{r}$, and $g(x_t)$ has the first absolute moment.

Assumption 3. *Let $E|x_{t_0}|^8 < \infty$. The discrete sample $\{x_{t_0+t\triangle_n}\}_{t=1}^n$ is observed at equal sampling intervals $\triangle_n = (T-t_0)/n$ and $\{x_{t_0+t\triangle_n}\}_{t=1}^n$ is strictly stationary and absolutely regular with the geometric decay rate.*

Assumption 4. *The parameter space, Θ, is compact. For any given $\theta \in \Theta$, $a_{ij}^0(x,\theta)$ satisfies Assumptions 1–2. $\hat{\theta}_n$ is a parameter estimator such that $\sqrt{n}(\hat{\theta}_n - \theta^*) = O_p(1)$, where $\theta^* \equiv plim\hat{\theta}_n$ is an interior element in Θ and $\theta^* = \theta_0$ under H_0.*

Assumption 5. *The kernel function $K(\cdot)$ is a product kernel of some univariate kernel $k(\cdot)$, i.e., $K(u) = \Pi_{j=1}^d k(u_j)$, where $k(\cdot) : R \to R^+$, which satisfies the Lipschitz condition and is a symmetric, bounded function with $\int k(s)ds = 1$. $|s|^d k(s) \to 0$ as $|s| \to \infty$, and $K(\cdot)$ is of the order r for the positive integer r.*

Assumption 6. *$w(x)$ is a given Borel measurable function and bounded with compact support, $S \subseteq D$. $f(x)$ is r times continuously differentiable. $f(x)$ and its derivatives are bounded on D, and $f(x)$ is bounded from zero on S. There exists $\alpha > 0$, such that $\int exp(\alpha \sum_{i=1}^{d} (x^i)^2) f(x) dx < \infty$.*

Assumption 1 ensures that the coefficients of the stochastic differential equation (1) are locally Lipschitz, under which a solution to (1) will be unique if it exists (strong uniqueness, Theorem 5.2.5 in Karatzas and Shreve, 1991). The global growth condition (17) in Assumption 2 ensures the existence of a solution to (1) by preventing the explosion of the process in finite time (Aït-Sahalia, 2008). Assumption 1, the global growth condition (17), and condition (18) in Assumption 2 will be used to provide useful upper bound inequality ((*A*.1) in the Appendix) on the higher even-order conditional moments of the solution. The upper-bound inequality will be used to prove the main asymptotic results in this paper. Assumption 3 requires that the underlying process $\{x_{n,t}, 1 \leq t \leq n\}$ be strictly stationary and absolutely regular with a geometric decay rate. The absolute regularity is required in order to use the central limit theorem for second-order degenerate U-statistics. $\{x_{n,t}, 1 \leq t \leq n\}$ is a strictly stationary process if positive integers s_1, s_2, and k, $\{x_{n,s_1}, x_{n,s_1+1}, \ldots, x_{n,s_1+k}\}$ and $\{x_{n,s_2}, x_{n,s_2+1}, \ldots, x_{n,s_2+k}\}$ have the same joint distribution, where $s_1 + k \leq n, s_2 + k \leq n$. The process $\{x_{n,t}, 1 \leq t \leq n\}$ is absolutely regular with the geometric decay rate if the mixing coefficient $\beta_{n,\tau} \equiv Sup_{s+\tau \leq n} E[Sup_{A \in M_{s+\tau,n}^{n}} \{P(A|M_{1,s}^{n}(n)) - P(A)\}] = O(\lambda^{\tau})$, where $M_{s,t}^{n}$ denotes the sigma algebra generated by $(x_{n,s}, \ldots, x_{n,t})$ for $s \leq t$ and λ is a positive constant, $0 < \lambda < 1$. Given that x_t is a stationary process, to prove absolute regularity with the geometric decay rate of the discrete observations from (1), it suffices to prove that the observed data sequence is geometrically ergodic (p.312, Mokkadem, 1988), which can be shown by the existence of a test function satisfying a Tweedie-type inequality (CD3, p. 536, Meyn and Tweedie, 1993). Following the approach provided by Meyn and Tweedie (pp.537–539, 1993), it can be shown that

the observed data sequence from the well-known class of multivariate affine diffusion processes satisfies the property of absolute regularity with the geometric decay rate.

In Assumption 4, we do not require any specific estimator for θ_0. Any $n^{1/2}$-consistent estimator can be used; for example, the minimum contract estimator in (7) proposed by Genon-Catalot and Jacod (1993), and the minimizing distance estimator in (8) proposed by Bandi and Phillips (2007). Assumption 5 imposes regularity conditions on the kernel function used in the nonparametric kernel estimations of the diffusion matrix and marginal density function of x_t. Under Assumption 5, Assumption 2 in Bandi and Moloche (2008), which is used to show the strong consistency and asymptotic normality of the nonparametric estimator of the diffusion matrix, is satisfied. Assumption 6 requires $w(x)$ to be bounded by compact support. This assumption will help us to avoid technical problems in proving the uniform convergence of the nonparametric estimations of the marginal density function, diffusion matrix, and their derivatives on S. In practice, S can typically be taken as the indicator function of a compact set related to the empirical question of interest.

With the above assumptions, the following theorem states the asymptotic distribution of the proposed test statistic J_{nij} under H_0^{ij}, and the asymptotic power of J_{nij} under the alternative H_a^{ij}.

Theorem 1. *Suppose* $h_n = O(n^{-1/\gamma})$, *where* $\frac{3d}{2} < \gamma < \frac{d}{2} + 2r$ *for* $d \leq 3$, *and* $2d + 1 < \gamma < \frac{d}{2} + 2r$ *for* $d \geq 4$. *If* T *is either fixed or* $T \to \infty$ *but* $Th_n^{d/2} \to 0$ *and* $\triangle_n \to 0$ *as* $n, T \to \infty$, *then under Assumptions 1–6, for* $1 \leq i \leq j \leq d$, *we have,*

(i) under $H_0, J_{nij} \to N(0,1)$ *in distribution as* $n \to \infty$, *and with* v_{nij}^2 *as a consistent estimator of* v^2, *where* $v^2 = 2\int (a_{ii}(x)a_{jj}(x) + a_{ij}^2(x))^2 f^4(x) w^2(x) dx \int [\int K(u)K(v+u)du]^2 dv$;

(ii) under $H_1, Pr(J_{nij} \geq B_n) \to 1$, *for any nonstochastic sequence* $B_n = o(nh_n^{d/2})$.

Remark 1. Note that from $\frac{3d}{2} < \gamma < \frac{d}{2} + 2r$ in Theorem 1 there is no need to use a high-order kernel ($r > 2$) unless the dimensionality of the diffusion process is greater than or equal to 3.

Under the assumptions on the smoothing parameter, we particularly have $nh_n^d \to \infty, nh_n^{d/2+2r} \to 0$, as $n \to \infty$, and $r \geq m = [\frac{d}{2}] + 1$.

Remark 2. Compared with the test statistics suggested by Chen and Hong (2010), Thompson (2008), Hong and Li (2005) and Aït-Sahalia (1996), in which each of their test statistics is shown to follow an asymptotic distribution under a null hypothesis, since the time span of the data tends to infinity, the asymptotic distributions of our test statistics under null hypotheses are derived by requiring the sampling frequency to increase; that is, our tests require the sampling interval $\Delta_n \longrightarrow 0$. The reason for the different requirements is that the marginal density function, the transition density, and the conditional characteristic function of a diffusion process can be identified by enlarging the sample size, even if the sample size increases, by letting the sampling frequency increase or prolonging the time period of the data. In our case, the identification of the diffusion matrix has to depend on the local dynamics of the diffusion process; that is, the evolution of the diffusion process for a small change in time.[9]

To test the null hypothesis H_0^{ij} versus H_A^{ij} at the level α, we need to compare J_{nij} to the critical value z_α from the $N(0,1)$ distribution, i.e., $z_{0.01} = 2.33$, $z_{0.05} = 1.64$, and $z_{0.10} = 1.28$ because the test J_{nij} is one-sided. We reject the null hypothesis when $J_{nij} > z_\alpha$.

The standard normal density function will be chosen as the kernel function in our simulation study and empirical applications. In practice, the choice of h_n is more important than the choice of $k(\cdot)$. Similar to Aït-Sahalia (2006), we choose $h_n = cn^{-1/\gamma}$, where $\gamma = 3.5$ for $d = 1$ and $\gamma = 5$ for $d = 3$, which satisfy the condition in Theorem 1. As with the choice of h_n in the estimation of the conditional moment in a nonparametric regression model, we choose c for the test statistic

[9]The diffusion function in the semiparametric diffusion process model of Aït-Sahalia (1996) can be identified by enlarging the observation period. This is because, given the parametric specification of the drift function in Aït-Sahalia's model, the perturbation caused by the diffusion function is actually implicitly identified. As a consequence, the diffusion function can be identified by enlarging the observation period.

J_{nij} by minimizing the estimated prediction error (EPE): $\text{EPE} = n^{-1}\sum_{t=1}^{n-1}\{[x_{n,t+1}^i - x_{n,t}^i][x_{n,t+1}^j - x_{n,t}^j]/\triangle_n - \hat{a}_{ij}(x_{n,t})\}^2$.

4 Finite Sample Performance

We now examine the finite sample performance of our tests using Monte Carlo simulations. For univariate diffusion processes, Monte Carlo simulations are reported in Li (2007), and the Monte Carlo simulation results suggest that the overall performance of the test is satisfactory. For multivariate diffusion processes, we focus on three-dimensional diffusion processes. Since the diffusion matrix is a symmetric matrix, we need only to evaluate the finite sample performance for our tests, $J_{n11}, J_{n12}, J_{n13}, J_{n22}, J_{n23}$ and J_{n33}. To examine the size performance, we simulate the data from a three-factor Vasicek model. Throughout this Monte Carlo simulation, the weighting function $w(x)$ is set as the indicator function of the set $S = [-2,2]\times[-2,2]\times[-2,2]$ in R^3.

DGP1. Three-factor Vasicek model:

$$d\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}\begin{pmatrix} \alpha_1 - x_{1t} \\ \alpha_2 - x_{2t} \\ \alpha_3 - x_{3t} \end{pmatrix} dt + \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix} d\begin{pmatrix} w_{1t} \\ w_{2t} \\ w_{3t} \end{pmatrix}. \quad (19)$$

The diffusion matrix of the three-factor Vasicek model is:

$$a(x_t) \equiv \{a_{ij}(x_t)\}_{1\le i,j\le 3} \equiv \begin{pmatrix} \sigma_{11}^2 & 0 & 0 \\ 0 & \sigma_{22}^2 & 0 \\ 0 & 0 & \sigma_{33}^2 \end{pmatrix}. \quad (20)$$

Given the diffusion matrix (20), the null hypotheses are: $a_{ii}(x_t) = \text{constant}$, for $i = 1,2,3$, $a_{1j}(x_t) = 0$ for $j = 2,3$, and $a_{23}(x_t) = 0$.

$\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$ in (20) is the long-run mean, and the eigenvalues of the matrix $\begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}$ determine the speed of the mean reversion of this diffusion process. The smaller the eigenvalues, the stronger the serial dependence in $\{x_t\}$ and, consequently, the slower the convergence to the long-run mean. As in the univariate diffusion case (Hong and Li, 2005), we are interested in the

impact of persistent dependence in $\{x_t\}$ on the size of the test statistics $J_{nij}, 1 \leq i \leq j \leq 3$. Following Aït-Sahalia and Kimmel (2010) and Hong and Li (2005), we set $(b_{11}, b_{21}, b_{22}, b_{31}, b_{32}, b_{33}, \sigma_{11}, \sigma_{22}, \sigma_{33}) = (0.50, -0.20, 1.00, 0.10, 0.20, 2.00, 1.00, 1.00, 1.00)$ and $(b_{12}, b_{13}, b_{23}, \alpha_1, \alpha_2, \alpha_3) = (0,0,0,0,0,0)$ for the low-persistence dependence case, while we set $(b_{11}, b_{21}, b_{22}, b_{31}, b_{32}, b_{33}, \sigma_{11}, \sigma_{22}, \sigma_{33}) = (0.20, -0.20, 0.5, 0.10, 0.20, 0.5, 1.00, 1.00, 1.00)$ and $(b_{12}, b_{13}, b_{23}, \alpha_1, \alpha_2, \alpha_3) = (0,0,0,0,0,0)$ for the case of high-persistence dependence.

The transition density of the vector of state variables is three-dimensional Gaussian. We simulate 1,000 data sets of the random sample $\{x^1_{n,t}, x^2_{n,t}, x^3_{n,t}\}_{t=1}^n$ at a monthly frequency for $n = 240, 480, 600$, and 840, respectively.[10] These sample sizes correspond to 20, 40, 50 and 70 years of monthly data. Given the null model DGP1, the minimum contrast estimators of parameters, $\sigma^2_{11}, \sigma^2_{22}$, and σ^2_{33}, are $\hat{\sigma}^2_{11} = \sum_{t=1}^{n-1}(x^1_{n,t+1} - x^1_{n,t})^2/T$, $\hat{\sigma}^2_{22} = \sum_{t=1}^{n-1}(x^2_{n,t+1} - x^2_{n,t})^2/T$, and $\hat{\sigma}^2_{22} = \sum_{t=1}^{n-1}(x^3_{n,t+1} - x^3_{n,t})^2/T$, respectively. In this simulation study, we set $T = 1$.[11]

Panel (a) and panel (b) in Table1 report the rejection rates of $J_{n11}, J_{n12}, J_{n13}, J_{n22}, J_{n23}$, and J_{n33} under DGP1 with low and high-persistence dependence, respectively, using the asymptotic critical values at the 1%, 5%, and 10% levels. Note that J_{n11}, J_{n22} and J_{n33} test for model misspecification in conditional variances (diagonal components) in the diffusion process, while J_{n12}, J_{n13}, and J_{n23} test for model misspecification in conditional covariances (off-diagonal components). Table 1 shows that the null hypothesis is either for a conditional variance or for a conditional covariance in the diffusion matrix, and that the corresponding test statistic J_{nij} $(1 \leq i \leq j \leq 3)$ has reasonable sizes at all three significance levels for sample sizes as small as $n = 240$. Note that the impact of persistent dependence on the size of J_{nij} is minimal, suggesting that J_{nij} $(1 \leq i \leq j \leq 3)$ can

[10] The initial values are drawn from the normal marginal density of the random sample. We discard the first 500 observations to eliminate any startup effects.

[11] To consider the impact of the time span on the test performance, we use different values of T to generate data; for instance, $T = 1, 5$, and $T = 10$. We only present the simulation results for $T = 10$. The simulation results for $T = 1$ and $T = 5$ are not presented because they are qualitatively similar to those for $T = 10$, but are available upon request.

achieve robustness to the persistent dependence. This result can be explained by the fact that the test statistics are independent of the specification of the drift matrix, which determines the level of the persistent dependence in the diffusion process.

To examine the power of J_{nij}, we stipulate that DGP1 is the null model, but the data are generated from the following four different affine diffusion processes (DGP2-DGP5), each of which captures the different direction of misspecification in the diffusion matrix.

DGP2. The three-factor affine diffusion process with time-varying conditional variances is:

$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \begin{pmatrix} \alpha_1 - x_{1t} \\ \alpha_2 - x_{2t} \\ \alpha_3 - x_{3t} \end{pmatrix} dt + \begin{pmatrix} x_{1t}^{1/2} & 0 & 0 \\ 0 & x_{2t}^{1/2} & 0 \\ 0 & 0 & x_{3t}^{1/2} \end{pmatrix} d \begin{pmatrix} w_{1t} \\ w_{2t} \\ w_{3t} \end{pmatrix}, \quad (21)$$

and its diffusion matrix is $a(x_t) \equiv \{a_{ij}(x_t)\}_{1\leq i,j\leq 3} = \begin{pmatrix} x_{1t} & 0 & 0 \\ 0 & x_{2t} & 0 \\ 0 & 0 & x_{3t} \end{pmatrix}$. The parameter values are taken from Aït-Sahalia and Kimmel (2010) and are set as $(b_{11}, b_{22}, b_{33}, \alpha_1, \alpha_2, \alpha_3) = (0.5, 2, 1, 2, 1, 1)$ and $(b_{12}, b_{13}, b_{21}, b_{23}, b_{31}, b_{32}) = (0, 0, 0, 0, 0, 0)$.

Given that DGP1 is the null model, DGP2 is used to examine the power performance of J_{nii} $(i = 1, 2, 3)$. Since the conditional covariances in DGP2 are correctly specified according to the null model DGP1, DGP2 is also used to examine the size performance of J_{nij} $(i \neq j)$ for conditional covariances when the conditional variances are misspecified.[12]

DGP3. The three-factor diffusion process with time-varying conditional variances and a conditional covariance$(a_{12}(x_t) \neq 0)$ is:

$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \begin{pmatrix} \alpha_1 - x_{1t} \\ \alpha_2 - x_{2t} \\ \alpha_3 - x_{3t} \end{pmatrix} dt + \begin{pmatrix} x_{1t}^{1/2} & x_{1t}^{1/2} & 0 \\ 0 & x_{2t}^{1/2} & 0 \\ 0 & 0 & x_{3t}^{1/2} \end{pmatrix} d \begin{pmatrix} w_{1t} \\ w_{2t} \\ w_{3t} \end{pmatrix}, \quad (22)$$

and its diffusion matrix is $a(x_t) \equiv \{a_{ij}(x_t)\}_{1\leq i,j\leq 3} = \begin{pmatrix} 2x_{1t} & \sqrt{x_{1t}x_{2t}} & 0 \\ \sqrt{x_{1t}x_{2t}} & x_{2t} & 0 \\ 0 & 0 & x_{3t} \end{pmatrix}$. The same val-

[12] Under DGP1, the on-diagonal components of DGP2 are misspecified, but the off-diagonal components are specified correctly. As a consequence, DGP2 is not only used to examine the power performance of J_{nii}, but also to examine the size performance of J_{nij} for $i \neq j$.

ues for the parameters are taken as in DGP2.

Although DGP3 has the same specification of the conditional variance $a_{ii}(x_t)$ as DGP 2, the conditional covariance $a_{12}(x_t)$ in DGP3 is misspecified. Consequently, DGP3 is used to examine the power performance of J_{n12} to detect the misspecification from the conditional covariance $a_{12}(x_t)$, while it is also used to examine the power performance of J_{nii}, for $i = 1,2,3$, when there is a misspecification in the conditional covariance $a_{12}(x_t)$.

In addition, given the null model DGP1, since DGP3 has the correctly specified conditional covariances $a_{13}(x_t)$ $(=0)$ and $a_{23}(x_t)$ $(=0)$, DGP3 is also used to investigate the size performance of J_{n13} and J_{n23} when there is a misspecification in both the conditional variances and the conditional covariance.

DGP4. The three-factor affine diffusion process with time-varying conditional variances and the conditional covariance $(a_{13}(x_t) \neq 0)$ is:

$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} dt + \begin{pmatrix} x_{1t}^{1/2} & 0 & x_{1t}^{1/2} \\ 0 & x_{2t}^{1/2} & 0 \\ 0 & 0 & x_{3t}^{1/2} \end{pmatrix} d \begin{pmatrix} w_{1t} \\ w_{2t} \\ w_{3t} \end{pmatrix}, \tag{23}$$

and its diffusion matrix is $a(x_t) \equiv \{a_{ij}(x_t)\}_{1 \leq i,j \leq 3} = \begin{pmatrix} 2x_{1t} & 0 & \sqrt{x_{1t}x_{3t}} \\ 0 & x_{2t} & 0 \\ \sqrt{x_{1t}x_{3t}} & 0 & x_{3t} \end{pmatrix}$. The same values for the parameters are taken as in DGP2.

Compared with DGP2 and DGP3, DGP4 is used to examine the power performance of J_{n13} to detect the misspecification from the conditional covariance $a_{13}(x_t)$, and is also used to examine the power performance of J_{nii}, for $i = 1,2,3$, when there exists a misspecification in the conditional covariance $a_{13}(x_t)$.

Since the conditional covariances $a_{12}(x_t)$ $(=0)$ and $a_{23}(x_t)$ $(=0)$ in DGP4 are correctly specified, DGP4 is also used to investigate the size performance of J_{n12} and J_{n23} when there exists a misspecification in both the conditional variances and the conditional covariance.

DGP5. The three-factor diffusion process with a time-varying conditional variance and conditional covariance($a_{23}(x_t) \neq 0$) is:

$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} dt + \begin{pmatrix} x_{1t}^{1/2} & 0 & 0 \\ 0 & x_{2t}^{1/2} & x_{2t}^{1/2} \\ 0 & 0 & x_{3t}^{1/2} \end{pmatrix} d \begin{pmatrix} w_{1t} \\ w_{2t} \\ w_{3t} \end{pmatrix}, \tag{24}$$

and its diffusion matrix is $a(x_t) \equiv \{a_{ij}(x_t)\} = \begin{pmatrix} x_{1t} & 0 & 0 \\ 0 & 2x_{2t} & \sqrt{x_{2t}x_{3t}} \\ 0 & \sqrt{x_{2t}x_{3t}} & x_{3t} \end{pmatrix}$. The same values for the parameters are taken as in DGP2.

DGP5 is used to examine the power performance of J_{n23} to detect the misspecification from the conditional covariance $a_{23}(x_t)$; it is also used to examine the power performance of J_{nii}, for $i = 1,2,3$, when there exists a misspecification in the conditional covariance $a_{23}(x_t)$.

Given that the conditional covariances $a_{12}(x_t)$ $(= 0)$ and $a_{13}(x_t)$ $(= 0)$ in DGP5 are correctly specified, DGP5 is also used to investigate the size performance of J_{n12} and J_{n13} when there exists a misspecification in the conditional variances and the conditional covariance.

Since the closed-form transition densities are not available for these alternative models, we use Milstein's scheme to simulate 1,000 data sets of the random sample $\{(x_{n,t}^1, x_{n,t}^2, x_{n,t}^3)\}_{t=1}^n$ for sample sizes $n = 240, 420, 600$, and 840, respectively. For each data set, we estimate the Vasicek model in (19) by the minimum contrast estimator.

Panel(a) and panel(b) in Table 2 report the simulation results from DGP2 and DGP3, respectively. Panel(a) shows that the estimated powers of J_{n11}, J_{n22} and J_{n33} increase with the sample size n and almost reach unity when $n = 840$, in line with the consistency property of these tests, suggesting that J_{n11}, J_{n22}, and J_{n33} have good power for detecting a misspecification of the conditional variances in DGP2 against the null model DGP1. Compared with the results in Table 1, panel(a) shows that J_{n12}, J_{n13}, and J_{n23} still have good size performance even when there exists a misspecification from the conditional variances $(a_{ii}(x_t), i = 1,2,3)$, indicating that the size performances of

J_{n12}, J_{n13}, and J_{n23} are not impacted by either the correct specification or the misspecification of the other components in the diffusion process. Panel(b) shows that J_{n12} has good power in detecting a misspecification due to the covariance $a_{12}(x_t)$, while J_{n11}, J_{n22}, and J_{n33} still maintain a good power performance despite the misspecified covariance $a_{12}(x_t)$. Furthermore, panel(b) shows that J_{13} and J_{23} have a good size performance even if both the conditional variances $(a_{ii}(x_t), i = 1,2,3)$ and the conditional covariance $a_{12}(x_t)$ are misspecified.

The simulation results from DGP4 and DGP5 are reported in panel(a) and panel(b) of Table 3, respectively. Panel(a) and panel(b) show that J_{n13} and J_{n23} are quite powerful against the misspecifications from $a_{13}(x_t)$ and $a_{23}(x_t)$, respectively. We observe that the power patterns of J_{n11}, J_{n22}, and J_{n33} under DGP4 and DGP5 against DGP1 are very similar to those under DGP2 and DGP3 against DGP1. The size performance of J_{n12} and J_{n23} in panel(a) and J_{n12} and J_{n13} in panel (b) is satisfactory in detecting misspecifications from the conditional covariances.

Overall, our simulation study shows that our test statistics $J_{nij}, 1 \leq i \leq j \leq 3$ have a satisfactory size performance even for highly persistent dependent data. In particular, our tests still have a satisfactory size performance even if there exists a misspecification in the other components of the diffusion matrix. Our tests perform rather well in detecting misspecifications coming from either the conditional variance or the conditional covariance.

5 Empirical Application

We now apply our tests to the diffusion functions in five popular univariate diffusion models for the daily seven-day eurodollar deposit spot rate used in Aït-Sahalia (1996) and Hong and Li (2005), and the diffusion matrix in multivariate affine term-structure models for monthly U.S. Treasury yields used in Duffee (2002) and Hong and Li (2005). Although these data have been used in ex-

tensive analyses of the term structure of interest rates (e.g., Hong and Li, 2005), how to use these data to directly detect the specification of the diffusion function or the diffusion matrix in a diffusion process remains an unanswered question. We reuse these data to examine the specifications of the diffusion functions and diffusion matrices and to evaluate their performance, which is of crucial importance in building up a model for the term structure of interest rates because volatility plays an important role in capturing the stochastic behavior of interest rate dynamics.

5.1 Scalar term-structure models of the spot interest rate

We first apply our test to the diffusion functions in five popular models of short-term interest rates considered in this literature; namely, the Vasicek model (1977): $dx_t = \kappa(\alpha - x_t)dt + \sigma dw_t$; the Cox, Ingersoll, and Ross model (CIR)(1985): $dx_t = \kappa(\alpha - x_t)dt + \sigma\sqrt{x_t}dw_t$; the Ahn and Gao model (1999): $dx_t = x_t[\kappa - (\sigma^2 - \kappa\alpha)x_t]dt + \sigma x_t^{3/2}dw_t$; the Chan, Karolyi, Longstaff, and Sanders model(CKLS)(1992): $dx_t = \kappa(\alpha - x_t)dt + \sigma x_t^{\rho}dw_t$; and the Aït-Sahalia nonlinear drift model (1996): $dx_t = (\alpha_1 x_t^{-1} + \alpha_0 + \alpha_1 x_t + \alpha_2 x_t^2)dt + \sqrt{\beta_0 + \beta_1 x_t + \beta_2 x_t^{\beta_3}}dw_t$.

We use the same data set as in Aït-Sahalia (1996) and Hong and Li (2005): the seven-day eurodollar deposit rate with 5,505 daily observations from 1 June 1973 to 25 February 1995. The detailed descriptive statistics of the data are provided in Aït-Sahalia (2006). Both the Aït-Sahalia (1996) test and the Hong and Li (2005) test firmly reject all five models, but their tests cannot provide evidence whether the diffusion functions in these models are misspecified. For each of the five popular models, suppose that we do not have any information on the functional form of the drift function, and our focus is on testing the parametric specification of the diffusion function in the model. The parameters in the diffusion functions are estimated by the minimum contrast method (7). In this empirical application, we use the univariate normal kernel function, and the smoothing parameter is determined by minimizing the EPE (page 14). The weighting function

$w(x)$ is chosen as the indicator function of the interval $S = [0, 0.5]$.

Table 4 reports the parameter estimations. Table 5 reports the test statistic values for the five models. The results show that the diffusion functions in all five models are overwhelmingly rejected at conventional significance levels, which suggests that the diffusion functions in these models are severely misspecified. The diffusion function in the Vasicek model has the worst performance, with the test statistic value 849.79, followed by the diffusion function in the CIR model, with 589.69. The diffusion function in the Ahn and Gao (1999) model dramatically reduces the test statistic value to 108.59, and the performance is further improved by the diffusion function in the Aït-Sahalia nonlinear drift model (1996), with 96.75, and the diffusion function in the CKLS model, with 39.36, which performs the best.

While Aït-Sahalia (1996) shows that the Aït-Sahalia nonlinear drift model outperforms the popular constant elasticity of variance (CEV) models in modeling the daily seven-day eurodollar spot interest rate, we find that the diffusion function in the CKLS model (1999) (with the CEV diffusion function) performs better than the diffusion function in the Aït-Sahalia nonlinear drift model. The difference is that the Aït-Sahalia test (1996) evaluates the performance of the density function implied by the model, while our test evaluates the performance of the diffusion function in the model. Therefore, although the Aït-Sahalia nonlinear drift model has better performance in fitting the density function than the CKLS model, it does not mean that the diffusion function in the Aït-Sahalia nonlinear drift model has better performance than the diffusion function in the CKLS model in fitting the true diffusion function. This result suggests that the specification of the drift is important for capturing interest rate dynamics, and confirms the findings from Aït-Sahalia (1996) and Stanton (1997) that the nonlinear specification of the drift improves the performance of the density function in modeling interest rate dynamics, but in contrast to the findings by Durham

(2003) who finds that allowing for additional flexibility beyond a constant term in the drift provides negligible benefits.[13]

The testing results indicate that none of the diffusion functions in the five univariate diffusion models adequately captures the dynamics of volatility, but some diffusion functions perform relatively better than others. To show the possible reasons for the rejection of the diffusion functions in the univariate interest rate models, Figure 1 plots the estimated nonparametric diffusion function and the parametric diffusion functions for different models. The first noticeable aspect of Figure1 is that the nonparametric diffusion function is an increasing function of the interest rate between 0 and 0.17. Above 0.17, however, the diffusion function slightly decreases and then increases quickly, which provides strong evidence to reject the constant or flat specification of the diffusion function in the Vasicek model (1977). The nonparametric diffusion function looks neither linear nor uniformly increasing, which provides strong evidence to reject the uniformly increasing pattern specified by the popular CEV models, such as the diffusion functions in the models of Cox-Ingersoll-Ross (1985), Ahn and Gao (1999), and CKLS (1992). In the Aït-Sahalia nonlinear drift model (1996), the diffusion function at low interest rates shows more volatility than that in the CKLS model, which is the major reason why the diffusion function in the Aït-Sahalia nonlinear drift model (1996) has a worse performance than that in the CKLS model.

5.2 Multivariate affine term-structure models

A multivariate affine term-structure model typically specifies that the instantaneous riskless rate r_t is an affine function of N latent state variables $x_t = (x_t^1, ..., x_t^N)'$,

$$r_t = \delta_0 + \delta' x_t, \tag{25}$$

[13] Our findings are consistent with the findings in Hong and Li (2005), who develop an omnibus nonparametric specification test for a diffusion process based on the transition density, which, unlike the marginal density function used by Aït-Sahalia (1996), captures the full dynamics of a diffusion process.

where δ_0 is a scalar and δ is an $N \times 1$ vector, and the state vector x_t follows an affine diffusion process,

$$dx_t = \kappa(\theta - x_t)dt + \Sigma\sqrt{S(x_t, \alpha, \beta)}dw_t, \tag{26}$$

where κ and Σ are $N \times N$ matrices, and $S(x_t, \alpha, \beta)$ is the diagonal matrix with elements

$$S_{ii} = \alpha_i + \beta_i' x_t, \tag{27}$$

with α_i a scalar and β_i an $N \times 1$ vector. Without loss of generality, the matrix Σ is supposed to be the identity matrix. This is because if Σ is not equal to identity, we can construct a new set of state variables $z_t \equiv \Sigma^{-1} x_t$, which is a diffusion process with a diagonal diffusion matrix (Aït-Sahalia and Kimmel, 2010).

Denote the times to maturity of the yields observed without error as $\tau_1, ..., \tau_N$. Under (25)–(27), it can be shown that the observed values of yields on zero coupon bonds $(y(x_t, \tau_1), ..., y(x_t, \tau_N))'$ can be expressed as an affine function of the state variables,

$$y_t \equiv \begin{pmatrix} y(x_t, \tau_1) \\ \vdots \\ y(x_t, \tau_N) \end{pmatrix} \equiv \begin{pmatrix} y_t^1 \\ \vdots \\ y_t^N \end{pmatrix} = \begin{pmatrix} \gamma_0(\tau_1) \\ \vdots \\ \gamma_0(\tau_N) \end{pmatrix} + \begin{pmatrix} \gamma(\tau_1)' \\ \vdots \\ \gamma(\tau_N)' \end{pmatrix} \begin{pmatrix} x_t^1 \\ \vdots \\ x_t^N \end{pmatrix}, \tag{28}$$

where the scalar function $\gamma_0(\tau)$ and the $N \times 1$ vector-valued function $\gamma(\tau)$ solve a pair of ordinary differential equations given in Dai and Singleton (2000). Equation (28) can be expressed in matrix form,

$$y_t = \Gamma_0 + \Gamma' x_t. \tag{29}$$

We first apply our tests to examine whether the affine specification of the state variable x_t can capture the volatility dynamics of U.S. Treasury yields y_t. Using Ito's lemma to y_t, we have

$$dy_t = \Gamma'\kappa\{[\Gamma']^{-1}\Gamma_0 + \theta - [\Gamma']^{-1} y_t\}dt + \Gamma'\sqrt{S([\Gamma']^{-1}[y_t - \Gamma_0], \alpha, \beta)}dw_t. \tag{30}$$

Equation (30) indicates that under (25)–(27), yield y_t follows an affine diffusion process. We empirically test whether the diffusion matrix of y_t follows an affine specification.

We use the same data as Duffee (2002) and Hong and Li (2005): monthly yields on zero-coupon bonds with 6-month, 2-year and 10-year maturities from January 1952 to December 1998. The zero-coupon bond yields are interpolated from coupon bond prices using the method in McCulloch and Kwon (1993), whose sample is extended by Bliss (1997) beyond February 1991. A time-series plot of the data is provided in Figure 2. Due to the shift in monetary policy, the years from 1980 to 1982 are characterized by substantially higher interest rate levels than the rest of the sample period. Table 6 reports the descriptive statistics. Dickey-Fuller non-stationarity tests have been conducted, and the presence of a unit root is rejected under the 5% significance level for each of the three yields.

We suppose that the diffusion matrix in the diffusion process of y_t is as follows,

$$a^0(y_t,\theta) \equiv \begin{pmatrix} a^0_{11}+\sum_{l=1}^3 a^0_{11l}y_t^l & \cdots & \cdots \\ a^0_{21}+\sum_{l=1}^3 a^0_{21l}y_t^l & a^0_{22}+\sum_{l=1}^3 a^0_{22l}y_t^l & \cdots \\ a^0_{31}+\sum_{l=1}^3 a^0_{31l}y_t^l & a^0_{32}+\sum_{l=1}^3 a^0_{32l}y_t^l & a^0_{33}+\sum_{l=1}^3 a^0_{33l}y_t^l \end{pmatrix}, \tag{31}$$

and that the true diffusion matrix of y_t is $a(y) = \{a_{ij}(y)\}_{1\le i,j\le 3}$. The null hypotheses we are interested in are :

$$a_{ij}(y_t) = a^0_{ij} + \sum_{l=1}^3 a^0_{ijl}y_t^l, 1 \le i \le j \le 3. \tag{32}$$

In the empirical application, we use the three-dimensional standard normal kernel function, and compute the smoothing parameter by minimizing EPE (page 14). The weighting function is chosen as the indicator function of the interval $S = \{(x,y,z)|x \in [0,0.50], y \in [0,0.50], z \in [0,0.50]\}$. We estimate the model parameters by the methods from the minimizing distance estimator (Eq.(8)). The parameter estimations and the test statistic of each component in the diffusion matrix $a^0(y_t,\theta_0)$ are reported in Table 7. Compared with upper-tailed $N(0,1)$ critical values (e.g., 2.33 at the 1%

level), the statistic J_{nij} strongly rejects the affine specification of each component in the diffusion matrix of the diffusion process of y_t, indicating that the linear specification of each component in the diffusion matrix of y_t cannot adequately capture the volatility dynamics in the observed data of y_t, and particularly implying that the joint linear specification of the drift vector and diffusion matrix in the diffusion process x_t is overwhelmingly rejected by the data.

To display possible reasons for the rejection of each component in an affine specification of the diffusion matrix of y_t, the nonparametric estimator of each component is reported in Figures 3 to 8. Given the nonparametric shape of each component in the diffusion matrix, the rejection is not surprising. The joint linear specification of the drift vector and diffusion matrix in x_t constrains each component of the diffusion matrix in y_t to be linearly increasing, while the relative nonparametric estimator looks neither increasing nor decreasing, i.e., it is nonmonotonic. The empirical evidence indicates that the misspecification of the diffusion matrix of y_t is caused jointly by the linear specification of the drift vector and the diffusion matrix in x_t.

Next, we use the test statistics to examine the performance of the popular three-factor affine term-structure models in Dai and Singleton (2000), $A_0(3), A_1(3), A_2(3)$, and $A_3(3)$, in capturing the volatility dynamics of y_t.[14] Hong and Li (2005) use their transition density-based tests to reject all these models, but their tests cannot be used to evaluate the performance of the diffusion matrix in an affine model in capturing the volatility dynamics of y_t. For each model, we estimate the parameters in equations (25) to (28) by Quasi-MLE, from which we derive the estimations of the parameters in the diffusion matrix in (30). Table 8 reports the test statistics for the components in the diffusion matrixes derived from $A_i(3), i = 0, 1, 2$, and 3. J_{nij} $(1 \leq i \leq j \leq 3)$ strongly rejects each component

[14]The three-factor affine term-structure models, $A_0(3), A_1(3), A_2(3)$, and $A_3(3)$, are defined in Dai and Singleton (2000) by taking different values of the parameters in equations (26) and (27). Γ_0 and Γ_1 in (29) can be obtained by solving numerically the ordinary differential equations with the initial conditions $\Gamma_0 = 0$ and $\Gamma = 0_{3\times 1}$ in Dai and Singleton (2000).

in the diffusion matrix in each model at conventional significance levels, although some models have better performance than others. Among the four affine models, $A_2(3)$ has the best overall performance across all components, while $A_0(3)$ performs the worst across all components. For $A_1(3)$ and $A_3(3)$, none of the models outperforms another across all components in the diffusion matrix.

Our empirical finding can be explained by the well-known trade-off between the flexibility in modeling the conditional variances of x_t and the conditional correlation between the components of x_t. $A_0(3)$ assumes that the conditional variances are constant, i.e., none of x_t affects the volatility of x_t, although it is more flexible in modeling the conditional correlation of x_t. $A_3(3)$ is characterized by the assumption that the volatility of each state variable is determined by the affine functions of x_t, but the requirements of admissibility (Dai and Singleton, 2000) impose strong restrictions on the correlations between the components of x_t. Dai and Singleton (2000) conjecture that the models that are able to accommodate both the time-varying volatilities of state variables and their time-varying correlations, such as $A_1(3)$ and $A_2(3)$, are more likely to perform better. Our empirical finding confirms the Dai and Singleton (2000) conjecture by showing that $A_2(3)$ has a better performance in capturing the volatility dynamics in modeling the 6-month, 2-year and 10-year maturities over the given time periods.

6 Conclusion

This paper focuses on testing the validity of the parametric specification of the diffusion matrix in a d-dimensional diffusion process. For this purpose, we test whether each component in this diffusion matrix is correctly specified. We propose a consistent test for the parametric specification of each component in the diffusion matrix without information on the functional form of the drift

vector. Consequently, not only can these tests detect whether the diffusion matrix is specified correctly, but they can also provide a clear indication of the direction of the misspecification.

We apply our tests to evaluate the performance of the diffusion functions of five popular univariate interest rates and the performance of the well-known multivariate affine term-structure models in capturing the volatility of the diffusion matrix of the diffusion process of yields. As a result, we obtain new empirical findings for the specification of the diffusion functions and diffusion matrixes in popular univariate term-structure models.

Table 1: **Percentage Rejection Rates of the True Null Hypothesis**

n	1%	5%	10%	1%	5%	10%	1%	5%	10%
Panel(a):	**Three-Factor Vasicek Model with Low Level of Persistent Dependence**								
	$H_0^{11}: a_{11}(x) = constant$			$H_0^{12}: a_{12}(x) = 0$			$H_0^{13}: a_{13}(x) = 0$		
	J_{n11}			J_{n12}			J_{n13}		
240	0.033	0.056	0.084	0.032	0.041	0.130	0.003	0.021	0.054
420	0.036	0.070	0.097	0.027	0.073	0.113	0.004	0.017	0.038
600	0.013	0.062	0.101	0.012	0.053	0.072	0.011	0.028	0.061
840	0.011	0.055	0.122	0.008	0.052	0.086	0.006	0.042	0.065
	$H_0^{22}: a_{22}(x) = constant$			$H_0^{23}: a_{23}(x) = 0$			$H_0^{33}: a_{33}(x) = constant$		
	J_{n22}			J_{n23}			J_{n33}		
240	0.017	0.042	0.069	0.024	0.052	0.092	0.024	0.045	0.092
420	0.021	0.055	0.063	0.019	0.060	0.073	0.015	0.058	0.132
600	0.032	0.048	0.062	0.021	0.064	0.121	0.011	0.061	0.078
840	0.008	0.054	0.078	0.013	0.052	0.071	0.009	0.062	0.092
Panel(b):	**Three-Factor Vasicek Model with High Level of Persistent Dependence**								
	$H_0^{11}: a_{11}(x) = constant$			$H_0^{12}: a_{12}(x) = 0$			$H_0^{13}: a_{13}(x) = 0$		
	J_{n11}			J_{n12}			J_{n13}		
240	0.025	0.062	0.111	0.031	0.047	0.121	0.004	0.034	0.075
420	0.030	0.061	0.120	0.025	0.057	0.082	0.007	0.044	0.066
600	0.019	0.041	0.121	0.022	0.059	0.087	0.014	0.053	0.072
840	0.012	0.047	0.108	0.008	0.048	0.091	0.012	0.054	0.096
	$H_0^{22}: a_{22}(x) = constant$			$H_0^{23}: a_{23}(x) = 0$			$H_0^{33}: a_{33}(x) = constant$		
	J_{n22}			J_{n23}			J_{n33}		
240	0.007	0.036	0.043	0.011	0.043	0.051	0.013	0.037	0.058
420	0.020	0.041	0.072	0.008	0.053	0.078	0.020	0.043	0.071
600	0.012	0.045	0.089	0.010	0.054	0.081	0.011	0.048	0.083
840	0.007	0.052	0.087	0.012	0.048	0.092	0.008	0.051	0.090

This table reports the empirical sizes of the test statistics, $J_{n11}, J_{n12}, J_{n13}, J_{n22}, J_{n23}$, and J_{n33}. The data are simulated from the three-factor Vasicek model (E.q. 19) at monthly frequencies for $n = 240, 420, 600$, and 840, respectively. For each data set, we estimate model parameters and compute test statistics J_{nij} $(1 \le i \le j \le 3)$. We set $(b_{11}, b_{21}, b_{22}, b_{31}, b_{32}, b_{33}, \sigma_{11}, \sigma_{22}, \sigma_{33}) = (0.50, -0.20, 1.00, 0.10, 0.20, 2.00, 1.00, 1.00, 1.00)$, and $(b_{12}, b_{13}, b_{23}, \alpha_1, \alpha_2, \alpha_3) = (0, 0, 0, 0, 0, 0)$ for the low-persistence dependence case; and $(b_{11}, b_{21}, b_{22}, b_{31}, b_{32}, b_{33}, \sigma_{11}, \sigma_{22}, \sigma_{33}) = (0.20, -0.20, 0.5, 0.10, 0.20, 0.5, 1.00, 1.00, 1.00)$, and $(b_{12}, b_{13}, b_{23}, \alpha_1, \alpha_2, \alpha_3)' = (0, 0, 0, 0, 0, 0)$ for the high-persistence dependence case.

Table 2: **Percentage Rejections of the False Null Hypothesis or the True Null Hypothesis**

n	1%	5%	10%	1%	5%	10%	1%	5%	10%
Panel(a): Three-factor affine diffusion process with time-varying conditional variance (DGP 2)									
	$H_0^{11}: a_{11}(x) = constant$			$H_0^{12}: a_{12}(x) = 0$			$H_0^{13}: a_{13}(x) = 0$		
	J_{n11}			J_{n12}			J_{n13}		
240	0.280	0.369	0.436	0.001	0.017	0.040	0.001	0.023	0.047
420	0.561	0.643	0.706	0.003	0.025	0.058	0.003	0.032	0.066
600	0.791	0.858	0.894	0.002	0.023	0.054	0.015	0.034	0.051
840	0.866	0.940	0.958	0.003	0.041	0.072	0.003	0.039	0.085
	$H_0^{22}: a_{22}(x) = constant$			$H_0^{23}: a_{23}(x) = 0$			$H_0^{33}: a_{33}(x) = constant$		
	J_{n22}			J_{n23}			J_{n33}		
240	0.211	0.275	0.323	0.001	0.011	0.032	0.301	0.372	0.482
420	0.398	0.402	0.451	0.013	0.025	0.042	0.411	0.456	0.523
600	0.710	0.814	0.883	0.003	0.034	0.040	0.753	0.834	0.901
840	0.908	0.921	0.943	0.005	0.036	0.042	0.924	0.930	0.943
Panel(b):Three-factor affine diffusion process with time-varying conditional covariance (DGP 3)									
	$H_0^{11}: a_{11}(x) = constant$			$H_0^{12}: a_{12}(x) = 0$			$H_0^{13}: a_{13}(x) = 0$		
	J_{n11}			J_{n12}			J_{n13}		
240	0.054	0.072	0.125	0.301	0.382	0.411	0.001	0.020	0.035
420	0.222	0.282	0.401	0.521	0.545	0.581	0.000	0.027	0.048
600	0.549	0.582	0.681	0.742	0.827	0.894	0.012	0.037	0.056
840	0.761	0.813	0.871	0.933	1.000	1.000	0.006	0.043	0.058
	$H_0^{11}: a_{22}(x) = constant$			$H_0^{12}: a_{23}(x) = 0$			$H_0^{13}: a_{33}(x) = constant$		
	J_{n22}			J_{n23}			J_{n33}		
240	0.291	0.305	0.361	0.005	0.013	0.040	0.142	0.177	0.218
420	0.413	0.446	0.517	0.021	0.043	0.057	0.319	0.341	0.407
600	0.756	0.832	0.911	0.005	0.034	0.059	0.662	0.716	0.743
840	0.833	0.902	0.957	0.013	0.065	0.084	0.805	0.834	0.922

For panel(a) in Table 2, the data are simulated from the three-factor affine diffusion process with time-varying conditional variances (E.q. 21) at a monthly frequency. Given that DGP1 is the null model, panel(a) reports the empirical powers of the test statistics, J_{11}, J_{22} and J_{33}, while panel(a) also reports the empirical sizes of the test statistics, J_{12}, J_{13}, and J_{23}. The parameters are set as $(b_{11}, b_{22}, b_{33}, \alpha_1, \alpha_2, \alpha 3) = (0.50, 2.00, 5.00, 2.00, 1.00)$ and $(b_{12}, b_{13}, b_{21}, b_{23}, b_{31}, b_{32}) = (0, 0, 0, 0, 0, 0)$. For panel(b) in Table 2, the data are simulated from the three-factor affine diffusion process with a time-varying conditional covariance (E.q. 22) at monthly frequencies. Given that DGP1 is the null model, panel(b) reports the empirical powers of the test statistics, J_{11}, J_{12}, J_{22}, and J_{33}, while panel (b) also reports the empirical sizes of J_{13} and J_{23}. The parameters are set as in DGP2.

Table 3: **Percentage Rejections of Either the False Null Hypothesis or the True Null Hypothesis**

n	1%	5%	10%	1%	5%	10%	1%	5%	10%
Panel(a):Three-factor affine diffusion process with time-varying conditional covariance (DGP 4)									
	$H_0^{11}: a_{11}(x) = constant$			$H_0^{12}: a_{12}(x) = 0$			$H_0^{13}: a_{13}(x) = 0$		
	J_{n11}			J_{n12}			J_{n13}		
240	0.201	0.254	0.439	0.023	0.061	0.142	0.159	0.236	0.358
420	0.421	0.583	0.617	0.007	0.038	0.096	0.297	0.356	0.477
600	0.654	0.767	0.809	0.004	0.046	0.115	0.479	0.560	0.733
840	0.913	0.928	0.947	0.006	0.043	0.087	0.809	0.921	0.933
	$H_0^{22}: a_{22}(x) = constant$			$H_0^{23}: a_{23}(x) = 0$			$H_0^{33}: a_{33}(x) = constant$		
	J_{n22}			J_{n23}			J_{n33}		
240	0.223	0.254	0.307	0.009	0.035	0.048	0.150	0.220	0.365
420	0.323	0.365	0.389	0.011	0.032	0.052	0.208	0.341	0.421
600	0.420	0.465	0.498	0.006	0.042	0.071	0.416	0.445	0.569
840	0.663	0.703	0.856	0.005	0.056	0.066	0.778	0.847	0.951
Panel(b):Three-factor affine diffusion process with time-varying conditional covariance (DGP 5)									
	$H_0^{11}: a_{11}(x) = constant$			$H_0^{12}: a_{12}(x) = 0$			$H_0^{13}: a_{13}(x) = 0$		
	J_{n11}			J_{n12}			J_{n13}		
240	0.179	0.205	0.258	0.001	0.022	0.026	0.000	0.007	0.009
420	0.311	0.355	0.401	0.002	0.030	0.041	0.005	0.012	0.055
600	0.457	0.498	0.502	0.006	0.025	0.039	0.006	0.034	0.060
840	0.781	0.835	0.901	0.012	0.030	0.041	0.007	0.043	0.075
	$H_0^{22}: a_{22}(x) = constant$			$H_0^{23}: a_{23}(x) = 0$			$H_0^{33}: a_{33}(x) = constant$		
	J_{n22}			J_{n23}			J_{n33}		
240	0.214	0.256	0.311	0.198	0.303	0.355	0.320	0.356	0.411
420	0.447	0.511	0.559	0.400	0.457	0.568	0.431	0.521	0.506
600	0.600	0.634	0.712	0.519	0.589	0.623	0.676	0.700	0.725
840	0.833	0.913	0.955	0.758	0.811	0.876	0.821	0.864	0.914

For panel(a) in Table 3, the data are simulated from the three-factor affine diffusion process with time-varying conditional variances (E.q. 23)at monthly frequencies. Given that DGP1 is the null model, panel(a) reports the empirical powers of the test statistics, J_{11}, J_{13}, J_{22}, and J_{33}, as well as the empirical sizes of the test statistics, J_{12} and J_{23}. The parameters are set as DGP 2. For panel (b) in Table 3, the data are simulated from the three-factor affine diffusion process with time-varying conditional covariances (E.q. 24) at monthly frequencies. Given that DGP1 is the null model, panel(b) reports the empirical powers of the test statistics, J_{11}, J_{22}, J_{23}, and J_{33}, as well as the empirical sizes of J_{12} and J_{13}. The parameters are set as DGP2.

Table 4: **The estimations of parameters in the diffusion functions of univariate models of interest rates**

Parameter	Vasicek	CIR	Ahn and Gao	CKLS	Nonlinear drift
β_0	0.0041	0	0	0	0.0089
	(0.0020)				(0.0012)
β_1	0	0.0527	0	0	-1.28
		(0.011)			(0.275)
β_2	0	0	3.196	1.47	1.67
			(0.561)	(0.185)	(0.0123)
β_3	0	0	3	2.57	1.13
				(0.746)	(0.465)

This table reports the estimations of parameters in the diffusion functions in five univariate diffusion models of spot interest rates using the seven-day eurodollar interest rates in Aït-Sahalia (1996) from 1 June 1973 to 25 February 1995. The diffusion function is generally specified as $\beta_0 + \beta_1 r + \beta_2 r^{\beta_3}$. Therefore, we have for the Vasicek model: diffusion function=β_0; for the CIR model: diffusion function=$\beta_1 r$; for the Ahn and Gao model: diffusion function=$\beta_2 r^3$; and for the nonlinear drift model: diffusion function=$\beta_0 + \beta_1 r + \beta_2 r^{\beta_3}$. Parameter estimations are obtained by the minimum contrast estimator method in Eq.(7). Standard errors are given in the parentheses.

Table 5: **Tests of the parametric specifications of the diffusion functions of univariate models of interest rates**

Diffusion function	Reference	Test statistic	Critical value	Result
β_0	Vasicek (1977)	849.79	1.65	reject
$\beta_1 x$	CIR (1985)	589.69	1.65	reject
$\beta_2 x^3$	Ahn and Gao (1999)	108.59	1.65	reject
$\beta_2 x^{\beta_3}$	CKLS (1992)	39.36	1.65	reject
$\beta_0 + \beta_1 x + \beta_2 x^{\beta_3}$	Aït-Sahalia (1996)	96.75	1.65	reject

This table reports the testing results for the diffusion functions in the five interest rate models using the seven-day eurodollar interest rates in Aït-Sahalia (1996) from June 1, 1973, to February 25, 1995. We have for Vasicek model: diffusion function=β_0; for CIR model: diffusion function=$\beta_1 r$; for Ahn and Gao model: diffusion fucntion= $\beta_2 r^3$; and for the nonlinear drift model: diffusion function=$\beta_0 + \beta_1 r + \beta_2 r^{\beta_3}$. Parameter estimations are obtained by the minimum contrast estimator method in Eq.(7).

Table 6: **Summary Statistics of the Data**

Frequency	Mean	Std.Dev	Skewness	Kurtosis	Min	Max	JB Test	A.D.F. Test
6-month	0.056	0.029	0.57	2.93	0.007	0.165	126.9	-2.66
2-year	0.061	0.028	0.81	3.64	0.012	0.161	71.87	-2.40
10-year	0.067	0.027	0.58	2.95	0.023	0.151	31.53	-1.71

JB test is the Jarque-Bera test, which tests for the normality of the unconditional distribution of yields. The full sample is from January 1952 to December 1998. The critical value of the Dickey-Fuller non-stationary test is -1.648 under the 5% significant level.

Table 7: **Testing the linear specification of the diffusion matrix in the diffusion process of yields**

Duffee's model (2002): monthly yields on zero-coupon bonds with 6-month, 2-year and 10-year maturities

$H_0^{11}: a_{11}(y_t) = a_{11}^0 + \sum_{l=1}^3 a_{11l}^0 y_t^l$

a_{10}^0	a_{111}^0	a_{112}^0	a_{113}^0
0.014	0.009	0.008	0.006
(0.003)	(0.001)	(0.000)	(0.000)

$J_{n11} = 34.27$

$H_0^{12}: a_{12}(y_t) = a_{12}^0 + \sum_{l=1}^3 a_{12l}^0 y_t^l$

a_{12}^0	a_{121}^0	a_{122}^0	a_{123}^0
0.011	0.007	0.006	0.004
(0.002)	0.002)	(0.000)	(0.000)

$J_{n12} = 39.17$

$H_0^{13}: a_{13}(y_t) = a_{13}^0 + \sum_{l=1}^3 a_{13l}^0 y_t^l$

a_{13}^0	a_{131}^0	a_{132}^0	a_{133}^0
0.005	0.003	0.003	0.002
(0.000)	(0.000)	(0.000)	(0.000)

$J_{n13} = 39.57$

$H_0^{22}: a_{22}(y_t) = a_{22}^0 + \sum_{l=1}^3 a_{22l}^0 y_t^l$

a_{22}^0	a_{221}^0	a_{222}^0	a_{223}^0
0.014	0.009	0.008	0.006
(0.003)	(0.001)	(0.000)	(0.000)

$J_{n22} = 69.56$

$H_0^{23}: a_{23}(y_t) = a_{23}^0 + \sum_{l=1}^3 a_{23l}^0 y_t^l$

a_{23}^0	a_{231}^0	a_{232}^0	a_{233}^0
0.011	0.007	0.006	0.004
(0.002)	0.002)	(0.000)	(0.000)

$J_{n23} = 64.95$

$H_0^{33}: a_{33}(y_t) = a_{33}^0 + \sum_{l=1}^3 a_{33l}^0 y_t^l$

a_{33}^0	a_{331}^0	a_{332}^0	a_{333}^0
0.005	0.003	0.003	0.002
(0.000)	(0.000)	(0.000)	(0.000)

$J_{n33} = 106.28$

This table reports the parameter estimations in the diffusion matrix of the diffusion process of yields and the test statistic values of $J_{nij}, 1 \le i \le j \le 3$ using monthly 6-month, 2-year and 10-year zero coupon yields from January 1952 to December 1998. Standard errors of the estimators of the parameters are given in the parentheses. The diffusion matrix has the affine specification:

$$a^0(y_t,\theta) \equiv \begin{pmatrix} a_{11}^0 + \sum_{l=1}^3 a_{11l}^0 y_t^l & \cdots & \cdots \\ a_{21}^0 + \sum_{l=1}^3 a_{21l}^0 y_t^l & a_{22}^0 + \sum_{l=1}^3 a_{22l}^0 y_t^l & \cdots \\ a_{31}^0 + \sum_{l=1}^3 a_{31l}^0 y_t^l & a_{32}^0 + \sum_{l=1}^3 a_{32l}^0 y_t^l & a_{33}^0 + \sum_{l=1}^3 a_{33l}^0 y_t^l \end{pmatrix},$$

Table 8: **The empirical performance of multivariate affine term-structure models**

	$a_{11}(y_t)$	$a_{12}(y_t)$	$a_{13}(y_t)$	$a_{22}(y_t)$	$a_{23}(y_t)$	$a_{33}(y_t)$
	J_{n11}	J_{n12}	J_{n13}	J_{n22}	J_{n23}	J_{n23}
$A_0(3)$	93.07	103.49	77.38	153.55	119.52	183.89
$A_1(3)$	41.16	46.86	47.90	88.14.71	89.92	288.57
$A_2(3)$	29.78	32.46	36.83	70.07	81.08	83.94
$A_3(3)$	45.15	40.26	48.49	92.67	108.37	217.42

This table reports the J_{nij} statistics for four affine models, $A_0(3), A_1(3), A_2(3)$, and $A_3(3)$ to capture the volatility dynamics of y_t using monthly 6-month, 2-year and 10-year zero-coupon Treasury yields from January 1952 to December 1998. For each model, $A_i(3), i = 0,1,2,3$, using Ito's Lemma to y_t, we have E.q.(30). We evaluate the performance of these models to capture the diffusion matrix of the diffusion process y_t.

Figure 1: Estimated Diffusion Functions for a Variety of Univariate Models of the Spot Interest Rate

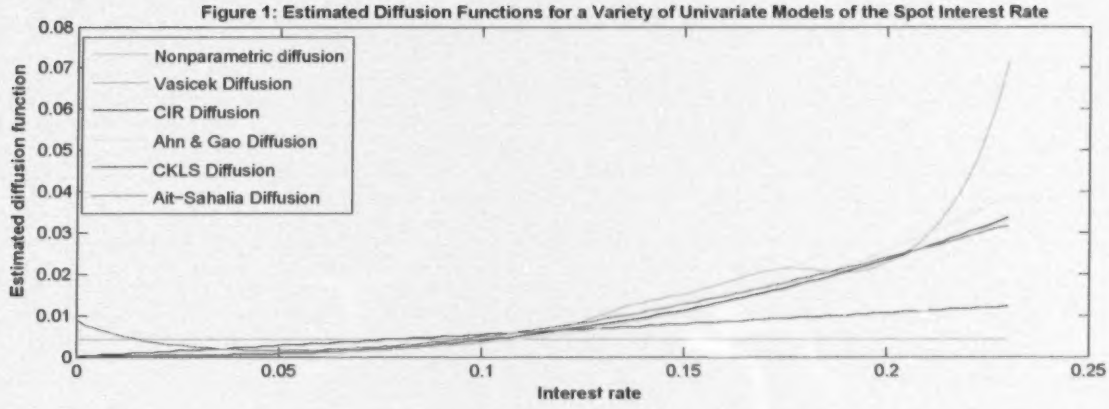


Figure 2. Monthly Yields on Zero-Coupon Bonds with 6-Month, 2- and 10-Year Maturities

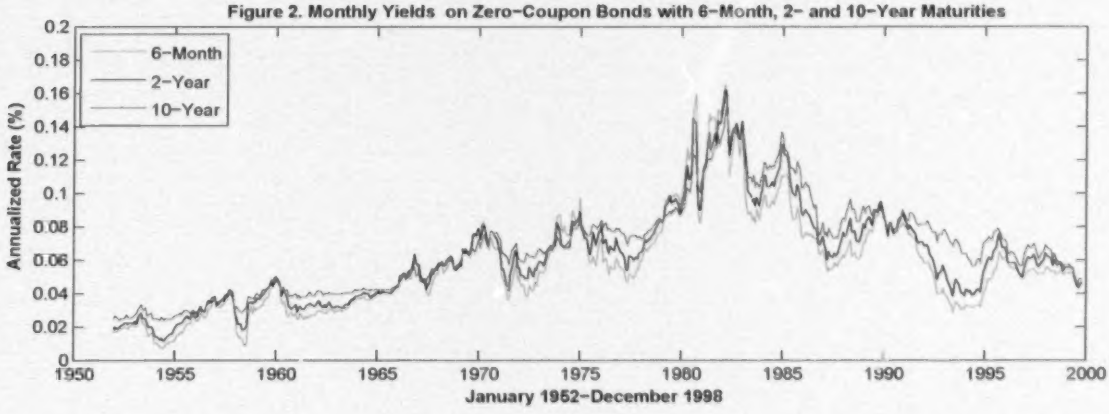


Figure 3. Nonparametric diffusion estimator of $a_{11}(y_1, y_2, 0.067)$

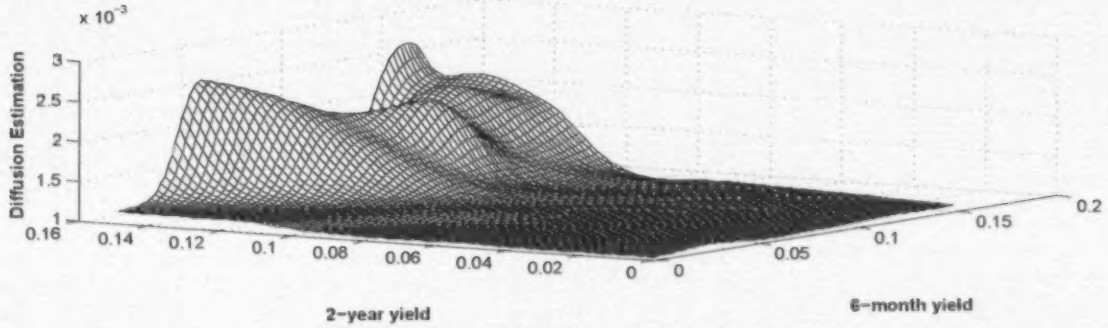


Figure 4. Nonparametric diffusion estimator of $a_{12}(y_1, y_2, 0.067)$

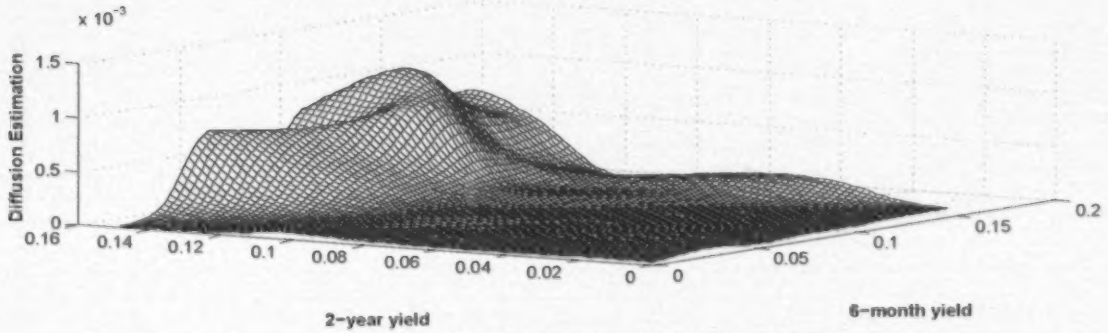


Figure 5. Nonparametric diffusion estimator of $a_{13}(y_1, 0.061, y_3)$

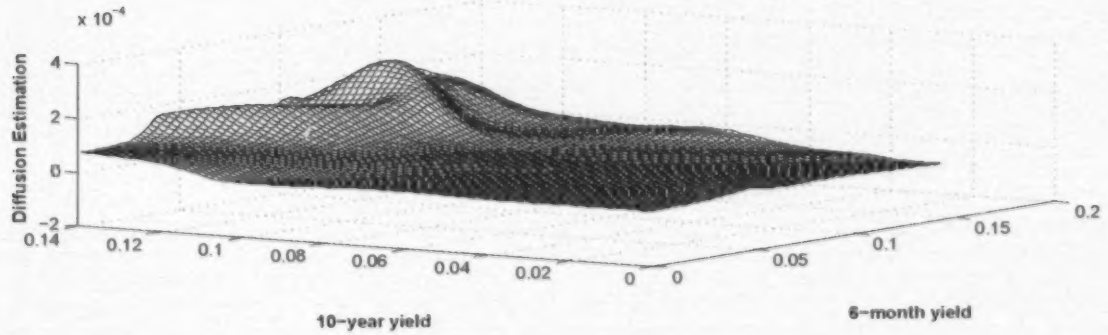


Figure 6. Nonparametric diffusion estimator of $a_{22}(0.061, y_2, y_3)$

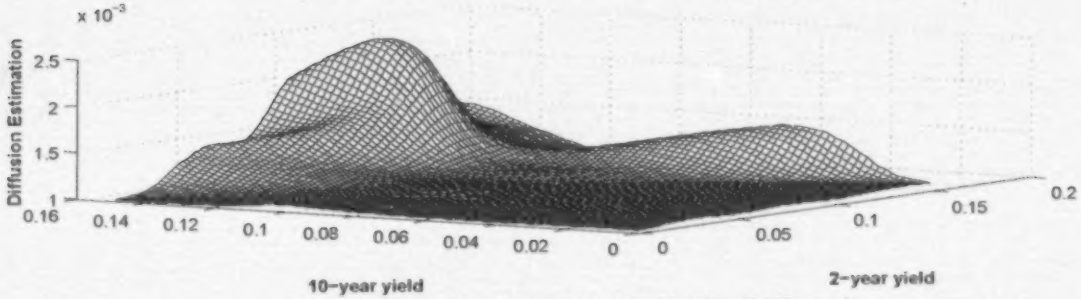


Figure 7. Nonparametric diffusion estimator of $a_{23}(0.061, y_2, y_3)$

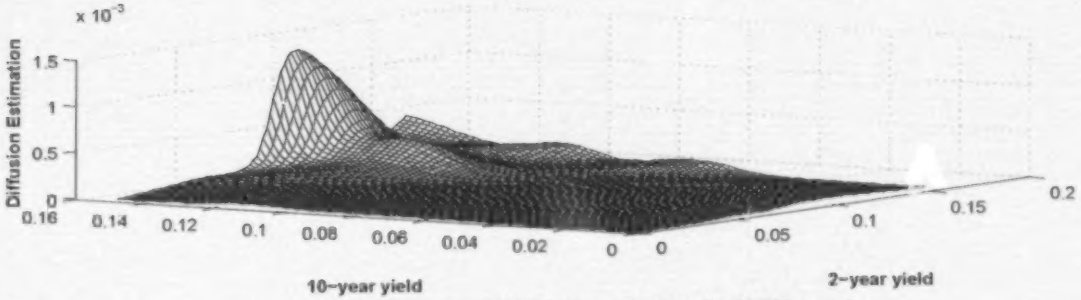
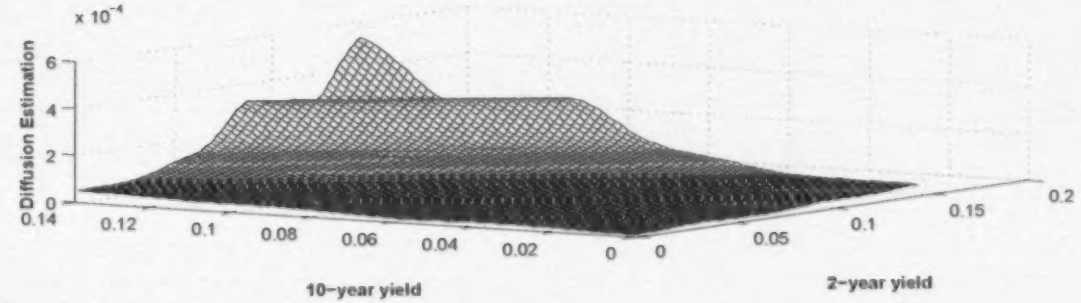


Figure 8. Nonparametric diffusion estimator of $a_{33}(0.061, y_2, y_3)$



References

Ahn, D., and B. Gao, 1999: "A Parametric Nonlinear Model of Term Structure Dynamics," , *Review of Financial Studies 12, 721-762.*

Aït-Sahalia, Y. (1996): " Testing Continuous-Time Models of the Spot Interest Rate," , *Review of Financial Studies 2, 385-426.*

Aït-Sahalia, Y., J.Fan, and H.Peng (2009): "Nonparametric Transition-Based Tests for Jump Diffusions," *Journal of the American Statistical Association, 104, 1102-1116.*

Aït-Sahalia, Y., P.J., Bickel, and T.M.Stoker (2001): "Goodness-of-Fit Tests for Kernel Regression with an Application to Option Implied Volatilities," *Journal of Econometrics, 105, 363-412.*

Aït-Sahalia, Y. (2008): "Closed-Form Likelihood Expansions for Multivariate Diffusions," *The Annals of Statistics, 36, No 2, 906-937.*

Aït-Sahalia, Y. and R.Kimmel (2010): "Estimating Affine Multifactor Term Structure Models Using Closed-Form Likelihood Expansions," *Journal of Financial Economics, 98, 113-144.*

Andersen, T.G., L.Benzoni and J.Lund (2002): " An Empirical Investigation of Continuous-Time Equity Return Models," *Journal of Finance, 57, 1239-1284.*

Bandi, F.M. and P. C. B. Phillips (2007): "A simple approach to the parametric estimation of potentially nonstationary diffusions, " *Journal of Econometrics, 137, 354-395*

Bandi, F.M. and G.Moloche (2008): "On the Functional Estimation of Multivariate Diffusion Processes," *Mimeo, Graduate School of Business at the University of Chicago.*

Bliss, R. (1997): "Testing Term Structure Estimation Methods," *Advances in Futures and Options Research, 9, 197-231.*

Chan, K.C., G.A. Karolyi, F.A. Longstaff, and A.B. Sanders (1992): "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate," *Journal of Finance, 47, 1209-1227.*

Chen, B. and Y. Hong (2010): "Characteristic Function-Based Testing for Multifactor Continuous-Time Markov Models Via Nonparametric Regression," *Econometric Theory, 26, 1115-1179.*

Chen, S. X., J.Gao, and C.Y.Tang (2008): "A Test for Model Specification of Diffusion Processes," *Annals of Statistics, 36, 167-198.*

Corradi, V. and N.R. Swanson (2005): "Bootstrap Specification Tests for Diffusion Processes, " *Journal of Econometrics, 124, 117-148.*

Cox, J.C., J.E. Ingersoll, and S.A.Ross, 1985: "A Theory of the Term Structure of Interest Rates," *Econometrica 53, 385-407.*

Dai, Q., and K.Singleton (2000): "Specification Analysis of Affine Term Structure Models," *Journal of Finance, 55, 1943-1978.*

Diebold, F.X., T. Gunther and A.S.Tay (1998): "Evaluating Density Forecasts, with Applications to Financial Risk Management," *International Economic Review, 39, 863-883.*

Duffee, G. (2002): " Term Premia and Interest Rate Forecasts in Affine Models, " *Journal of Finance, 57, 405-443.*

Durham, G.B. (2003): "Likelihood-Based Specification Analysis of Continuous-Time Models of the Short-Term Interest Rate, " *Journal of Financial Economics, 70, 43-487.*

Fan, Y. (1994): "Testing the Goodness-of-Fit of a Parametric Density Function by Kernel Method," *Econometric Theory, 10, 316-356.*

Fan, Y. and Q. Li (1999): "Central Limit Theorem for Degenerate U-Statistics of Absolutely Regular Processes with Applications to Model Specification Testing," *Nonparametric Statistics, 10, 245-271.*

Friedman, A. (1975): "Stochastic Differential Equations and Applicatins," *Vol. 1. Academic Press.*

Gallant, A.R. and G. Tauchen (1996): "Which Moments to Match?" *Econometric Theory, 12, 657-681.*

Genon-Catalot, V., and J.Jacod (1993): "On the Estimation of the Diffusion Coefficient for Multi-Dimensional Diffusion Processes," *Ann. Inst. Henri Poincaré: Probab. Stat., 29, 119-151.*

Györfi, L., W. Härdle, P. Sarda, and P.Vieu (1989): "Nonparametric Curve Estimation from Time Series. Lecture Notes in Statistics," *Vol. 60. Springer, New York.*

Härdle, W. and S.Luckhaus (1984): "Uniform Consistency of a Class of Regression Function Estimators," *Annals of Statistics, 12, 612-623.*

Hong,Y. and H.Li (2005): "Nonparametric Specification Testing for Continuous-Time Models with Application to Spot Interest Rates," *Review of Financial Studies, 18, 37-84.*

Hsin, C-W.(1995): "An Empirical Investigation of the Two-Factor Brennan-Schwartz Term Structure Model," *Review of Quantitative Finance and Accounting, 5, 71-92.*

Karatzas, I. and S.E.Shreve (1991): "Brownian Motion and Stochastic Calculus," *Springer, New York. MR1121940.*

Kloeden, P.E. and E.Platen (1992): "Numerical Solution of Stochastic Differential Equations," *Springer-Verlag Berlin Heidelberg.*

Li, F. (2007): "Testing the Parametric Specification of the Diffusion Function in a Diffusion Process," *Econometric Theory, 23, 221-250.*

Li, F. and G.Tkacz (2006): "A Consistent Bootstrap Test for Conditional Density Functions with Time-Series Data," *Journal of Econometrics, 133, 863-886.*

McCulloch, J., and H. Kwon (1993): "U.S. Term Structure Data, 1947-1991," *Ohi State University Working Paper, 93-96.*

Meyn, S., and R.L.Tweedie (1993): "Stability of Markovian Processes III: Forster-Lyapunov Criteria for Continuous-Time Processses," *Adv. Appl.Prob, 25, 518-548.*

Mokkadem, A. (1988): "Mixing Properties of Arma Processes (1988)," *Stochstic Processes and Their Application, 29, 309-315.*

Pagan, A. and A.Ullah (1999): "Nonparametric Econometrics," *Cambridge University Press.*

Piazzesi, M. (2010): "Affine Term Structure Models," *Handbook of Financial Econometrics, 691-766.* Edited by Yaciao Aït-Sahalia and Lars Peter Hansen, North Holland, Elsevier.

Stroock, D.W. and S.R.S. Varadhan (1979): "Multidimensional Diffusion Processes," *Springer, New York. MR0532498.*

Thompson, S. (2008): "Identifying Term Structure Volatility from the LIBOR-Swap Curve," *The Review of Financial Studies, V.21, 2, 819-854.*

Vasicek, O. (1977): "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics 5, 177-188.*

Yoshihara, K. (1976): "Limiting Behavior of Generalized Quadratic Forms Generated by Absolutely Regular Sequences II," *Z. Wahrscheinlichkeitstheorie verw. Gebiete., 35, 237-252.*

Appendix

Let $F(x)$ express the cumulative distribution function of x_t, and E^j the conditional expectation with respect to the σ-field generated by $\{x_u : u \le t_0 + j\triangle_n\}$. Also let $B(\cdot,\cdot,\cdot) : R^d \times R^d \times R^d \to R$ be a Borel measurable function and $F_n(x,\bar{x})$ be the joint distribution function for $(x_{n,i}, x_{n,j})$, where $i \neq j$. We denote $E_i[B(x_{n,i},x_{n,j},x_{n,k})] \equiv \int B(x,x_{n,j},x_{n,k})dF(x)$, $E_k[B(x_{n,i},x_{n,j},x_{n,k})] \equiv \int B(x_{n,i},x_{n,j},x)dF(x)$, and $E_{i,j}[B(x_{n,i},x_{n,j},x_{n,k})] \equiv \int B(x,\bar{x},x_{n,k})dF_n(x,\bar{x})$. h_n and $K((x_{n,j}-x)/h_n)$ are expressed by h and $K_j(x)$, respectively. The symbol C denotes a generic big enough positive constant. Recall $m = [\frac{d}{2}] + 1$.

Lemma 1. *Suppose that Assumptions 1-6 hold, $E(|x_{t_0}|^{2l}) < \infty$ for some positive integer l. Then,*
(i) for $t_0 + j\triangle_n < t' < T$,

$$E^j(|x_{t'} - x_{n,j}|^{2l}) \le D_n(1 + |x_{n,j}|^{2l})(t' - t_0 - j\triangle_n)^l, \tag{A.1}$$

where $D_n = 2^{2(2l-1)}C_D^{2l}e^{2l(2l+1)C_D^2(t'-t_0-j\triangle_n)}\{(t'-t_0-j\triangle_n)^l + [l(2l-1)]^l\}$;

(ii) let $s_n(z_{n,l},x,\theta) = (1/nh^d\triangle_n)K_l(x)[(x_{n,l+1}^j - x_{n,l}^j)(x_{n,l+1}^i - x_{n,l}^i) - \triangle_n a_{i,j}^0(x,\theta)]$, where $z_{n,l} = (x_l^i, x_{l+1}^i, x_l^j, x_{l+1}^j)$. Then, under the null hypothesis, for $x \in S$, we have,

$$E[s_n(z_{n,l},x,\theta_0)] = O(n^{-1}\triangle_n^{1/2}) + O(n^{-1}h^r), \tag{A.2}$$

and for $d \le 3$:

$$\begin{aligned}\int Es_n^2(z_{n,j},x,\theta_0)w(x)dF(x) &= (n^2h^d)^{-1}\int [a_{ii}(x)a_{jj}(x) + a_{ij}^2(x)]w(x)f(x)dF(x)\int K^2(u)du \\ &\quad + O(n^{-2}h^{-d+2}) + O(n^{-2}h^{-d}\triangle_n^{1/2}),\end{aligned} \tag{A.3}$$

for $d \ge 4$:

$$\begin{aligned}&\int Es_n^2(z_{n,j},x,\theta_0)w(x)dF(x) \\ &= (n^2h^d)^{-1}\int [a_{ii}(x)a_{jj}(x) + a_{ij}^2(x)]w(x)f(x)dF(x)\int K^2(u)du\end{aligned}$$

$$
\begin{aligned}
&+(n^2h^d)^{-1}\sum_{l=2}^{m-1}\frac{h^l}{l!}\int\int[(u'\nabla)^l(a_{ii}(x)a_{jj}(x)f(x))]w(x)K^2(u)dF(x)du\\
&+2(n^2h^d)^{-1}\sum_{l=2}^{m-1}\frac{h^l}{l!}\int\int[(u'\nabla)^l(a_{ij}^2(x)f(x))]w(x)K^2(u)dF(x)du\\
&-2(n^2h^d)^{-1}\sum_{l=2}^{m-1}\frac{h^l}{l!}\int\int[(u'\nabla)^l(a_{ij}(x)f(x))]a_{ij}(x)w(x)K^2(u)dF(x)du\\
&+(n^2h^d)^{-1}\sum_{l=2}^{m-1}\frac{h^l}{l!}\int\int[(u'\nabla)^l f(x)]a_{ij}^2(x)w(x)K^2(u)dF(x)du\\
&+O(n^{-2}h^{-d}\triangle_n^{1/2})+O(n^{-2}h^{-d+m}),
\end{aligned}
\tag{A.4}
$$

where $m=[d/2]+1$.

Proof of Lemma 1: we only prove (A.3) and (A.4) because (A.1) and (A.2) can be proven by following a method similar to the one for proving (A.2) in Lemma 2 in Li (2007).

For notational simplicity, we denote $\int_{t_0+j\triangle_n}^{t_0+(j+1)\triangle_n}G(u)du$ by $\int_{\triangle_n}G(u)du$, where $G(u)$ is any integrable function. To prove (A.3) and (A.4), applying the Itô formula to

$(x_{n,l+1}^i-x_{n,l}^i)^2(x_{n,l+1}^j-x_{n,l}^j)^2$, $(x_{n,l+1}^i-x_{n,l}^i)(x_{n,l+1}^j-x_{n,l}^j)$, $(x_{n,l+1}^j-x_{n,l}^j)^2$, and $(x_{n,l+1}^i-x_{n,l}^i)^2$, respectively, under H_0^{ij} we have,

$$
\begin{aligned}
&E\int[s_n(z_{n,l},x,\theta_0)]^2w(x)dF(x)\\
=\;&(nh^d\triangle_n)^{-2}E\{\int K_l^2(x)[(x_{n,l+1}^j-x_{n,l}^j)(x_{n,l+1}^i-x_{n,l}^i)-a_{i,j}^0(x)\triangle_n]^2w(x)dF(x)\}\\
=\;&(nh^d\triangle_n)^{-2}\{\int E\{K_l^2(x)\int_{\triangle_n}2[(x_u^i-x_{n,l}^i)(x_u^j-x_{n,l}^j)^2\mu_i(x_u)\\
&+(x_u^i-x_{n,l}^i)^2(x_u^j-x_{n,l}^j)\mu_j(x_u)]du\}w(x)dF(x)\\
&-2\triangle_n\int E\{K_l^2(x)\int_{\triangle_n}[(x_u^j-x_{n,l}^j)\mu_i(x_u)+(x_u^i-x_{n,l}^i)\mu_j(x_u)]du\}w(x)dF(x)\\
&+\int E\{K_l^2(x)\int_{\triangle_n}[4(x_u^j-x_{n,l}^j)(x_u^i-x_{n,l}^i)(a_{ij}(x_u)-a_{ij}(x_{n,l}))\\
&+(x_u^j-x_{n,l}^j)^2(a_{ii}(x_u)-a_{ii}(x_{n,l}))+(x_u^i-x_{n,l}^i)^2(a_{jj}(x_u)-a_{jj}(x_{n,l}))]du\}w(x)dF(x)\\
&+\int E\{K_l^2(x)\int_{\triangle_n}\int_{t_0+l\triangle_n}^{u}[4(x_s^j-x_{n,l}^j)\mu_i(x_s)+4(x_s^i-x_{n,l}^i)\mu_j(x_s)]a_{ij}(x_{n,l})dsdu\}w(x)dF(x)
\end{aligned}
$$

$$
\begin{aligned}
&+4\int E\{K_l^2(x)\int_{\triangle_n}\int_{t_0+\triangle_n}^{u}[a_{ij}(x_s)-a_{ij}(x_{n,l})]a_{ij}(x_{n,l})dsdu\}w(x)dF(x)\\
&+\int E\{K_l^2(x)\int_{\triangle_n}\int_{t_0+\triangle_n}^{u}[(a_{jj}(x_u)-a_{jj}(x_{n,l}))a_{ii}(x_{n,l})+(a_{ii}(x_u)-a_{ii}(x_{n,l}))a_{jj}(x_{n,l})]\\
&\times dsdu\}w(x)dF(x)\\
&+2\int E\{K_l^2(x)\int_{\triangle_n}\int_{t_0+\triangle_n}^{u}[(x_s^j-x_{n,l}^j)\mu_i(x_s)a_{jj}(x_{n,l})+(x_s^j-x_{n,l}^j)\mu_i(x_s)a_{ii}(x_{n,l})]\\
&\times dsdu\}w(x)dF(x)\\
&-2\triangle_n\int E\{K_l^2(x)\int_{\triangle_n}(a_{ij}(x_u)-a_{ij}(x_{n,l}))a_{ij}(x)duw(x)\}dF(x)\\
&+2\int E\{K_l^2(x)\int_{\triangle_n}\int_{t_0+\triangle_n}^{u}a_{ii}(x_{n,l})a_{jj}(x_{n,l})w(x)dsdu\}dF(x)\\
&+4\int E\{K_l^2(x)\int_{\triangle_n}\int_{t_0+\triangle_n}^{u}a_{ij}^2(x_{n,l})w(x)dsdu\}dF(x)\\
&-2\triangle_n\int E\{K_l^2(x)\int_{\triangle_n}a_{ij}(x_{n,l})a_{i,j}(x)w(x)du\}dF(x)\\
&+\triangle_n^2\int E\{K_l^2(x)a_{ij}^2(x)w(x)\}dF(x)\}\\
\equiv\ & B_n^1+\ldots+B_n^{12}. \qquad \text{(A.5)}
\end{aligned}
$$

From $(x_u^{i'}-x_{n,l}^{i'})^2\le|x_u-x_{n,l}|^2$, $i'=i$ or $i'=j$, Schwarz's inequality, (A.1), and $\triangle_n h^{-d}=\frac{Th^{d/2}}{n^{(2\gamma-d3)/2\gamma}}=o(1)$, we have $B_n^i=O(n^{-2}h^{-d}\triangle_n^{1/2})$ for $i=1,\ldots,8$. Now, we consider B_n^i for $i=9,\ldots,12$.

Using the changing variable $u=\frac{x_{n,l}-x}{h}$, we have,

$$
\begin{aligned}
B_n^9 &= (nh^d)^{-2}\int E\{K_l^2(x)a_{ii}(x_{n,l})a_{jj}(x_{n,l})w(x)\}dF(x)\\
&= (n^2h^d)^{-1}\int\int K^2(u)a_{ii}(x+hu)a_{jj}(x+hu)f(x+hu)w(x)dudF(x)\\
&= (n^2h^d)^{-1}\int a_{ii}(x)a_{jj}(x)f(x)w(x)dF(x)\int K^2(u)du\\
&\quad+(n^2h^d)^{-1}\frac{h^2}{2}\int\int[(u'\nabla)^2(a_{ii}(x)a_{jj}(x)f(x))]w(x)K^2(u)dF(x)du+\ldots\\
&\quad+(n^2h^d)^{-1}\frac{h^{m-1}}{(m-1)!}\int\int[(u'\nabla)^{m-1}(a_{ii}(x)a_{jj}(x)f(x))]w(x)K^2(u)dF(x)du\\
&\quad+O(n^{-2}h^{-d+m}), \qquad \text{(A.6)}
\end{aligned}
$$

where ∇ is the differential operator, which is defined on page 8. To obtain (A.6), we have used the Taylor expansion of $a_{ii}(x+hu)a_{jj}(x+hu)f(x+hu)$ at x. Similarly, we have,

$$\begin{aligned}
B_n^{10} &= 2(nh^d)^{-2}\int E\{K_l^2(x)a_{ij}^2(x_{n,l})\}w(x)dF(x) \\
&= 2(n^2h^d)^{-1}\int\int K^2(u)a_{ij}^2(x+hu)f(x+hu)duw(x)dF(x) \\
&= 2(n^2h^d)^{-1}\int a_{ij}^2(x)f(x)w(x)dF(x)\int K^2(u)du, \\
&\quad +(n^2h^d)^{-1}h^2\int\int[(u'\nabla)^2(a_{ij}^2(x)f(x))]w(x)K^2(u)dF(x)du+... \\
&\quad +2\frac{(n^2h^d)^{-1}h^{m-1}}{(m-1)!}\int\int[(u'\nabla)^{m-1}(a_{ij}^2(x)f(x))]w(x)K^2(u)dF(x)du \\
&\quad +O(n^{-2}h^{-d+m}), \qquad \text{(A.7)} \\
B_n^{11} &= -2(nh^d)^{-2}\int E\{K_l^2(x)a_{ij}(x_{n,l})\}a_{i,j}(x)w(x)dF(x) \\
&= -2(n^2h^d)^{-1}\int\int a_{ij}(x+hu)f(x+hu)a_{ij}(x)w(x)K^2(u)dF(x)du \\
&= -2(n^2h^d)^{-1}\int a_{ij}^2(x)f(x)w(x)dF(x)\int K^2(u)du \\
&\quad -(n^2h^d)^{-1}h^2\int\int[(u'\nabla)^2(a_{ij}(x)f(x))]a_{ij}(x)w(x)K^2(u)dF(x)du-... \\
&\quad -2(n^2h^d)^{-1}\frac{h^{m-1}}{(m-1)!}\int\int[(u'\nabla)^{m-1}(a_{ij}(x)f(x))]a_{ij}(x)w(x)K^2(u)dF(x)du \\
&\quad +O(n^{-2}h^{-d+m}), \qquad \text{(A.8)} \\
B_n^{12} &= (n^2h^d)^{-1}\int EK_l^2(x)a_{ij}^2(x)w(x)dF(x) \\
&= (n^2h^d)^{-1}\int\int a_{ij}^2(x)f(x+hu)K^2(u)w(x)dF(x)du \\
&= (n^2h^d)^{-1}\int a_{ij}^2(x)f(x)w(x)dF(x)\int K^2(u)du \\
&\quad +(n^2h^d)^{-1}\frac{h^2}{2}\int\int[(u'\nabla)^2(f(x))]a_{ij}^2(x)w(x)K^2(u)dF(x)du+... \\
&\quad +(n^2h^d)^{-1}\frac{h^{m-1}}{(m-1)!}\int\int[(u'\nabla)^{m-1}(f(x))]a_{ij}^2(x)w(x)K^2(u)dF(x)du \\
&\quad +O(n^{-2}h^{-d+m}). \qquad \text{(A.9)}
\end{aligned}$$

From $B_n^i = O(n^{-2}h^{-d})\triangle_n^{1/2}$ $(i=1,...,8)$ and (A.5) to (A.9), $(A.4)$ holds for any $d \geq 1$. However,

if $d \leq 3$, we have,

$$\begin{aligned} & B_{n1}^9 + B_{n1}^{10} + B_{n1}^{11} + B_{n1}^{12} \\ = \; & (n^2h^d)^{-1} \int\int (a_{ij}^2(x) + a_{ii}(x)a_{jj}(x)) f(x) w(x) K^2(u) dF(x) du \\ & + O(n^{-2}h^{-d+2}). \end{aligned} \tag{A.10}$$

$B_n^i = O(n^{-2}h^{-d}\triangle_n^{1/2})$ $(i = 1, ..., 8)$ and $(A.5) - (A.10)$ imply $(A.3)$ and $(A.4)$.

Lemma 2. *Under the same assumptions as in Theorem 1, we have,*

$$sup_{x \in S} |a_{ij}(x) - \hat{a}_{ij}(x)| = O_p(h^r + n^{-1/2}h^{-d/2} ln(n) + \triangle_n^{1/2}). \tag{A.11}$$

$$sup_{x \in S} |\frac{\partial^L \hat{a}_{ij}(x)}{\partial x^{d_1} ... \partial x^{d_L}} - \frac{\partial^L a_{ij}(x)}{\partial x^{d_1} ... \partial x^{d_L}}| = O_p(h^{r-L} + n^{-1/2}h^{-L-d/2} ln(n) + \triangle_n^{1/2} h^{-L}), \tag{A.12}$$

where $L \leq m-1$, and $1 \leq d_i \leq d$, for $i = 1, 2, ..., L$.

Proof of Lemma 2: to prove $(A.11)$, let $v_n(x) = n^{-1}h^{-d} \sum_{t=1}^{n-1} K(\frac{x_{n,t}-x}{h}) \triangle_n^{-1} [x_{n,t+1}^i - x_{n,t}^i][x_{n,t+1}^j - x_{n,t}^j]$ and $v(x) = a_{ij}(x) f(x)$, then we have the following decomposition,

$$\hat{a}_{ij}(x) - a_{ij}(x) = [S_1(x) + S_2(x) - a_{ij}(x) S_3(x) - a_{ij}(x) S_4(x)] (\hat{f}(x))^{-1}, \tag{A.13}$$

where $S_1(x) = v_n(x) - Ev_n(x), S_2(x) = Ev_n(x) - v(x), S_3(x) = \hat{f}(x) - E\hat{f}(x)$, and $S_4(x) = E\hat{f}(x) - f(x)$. Since $f(x)$ is bounded away from zero on S, it is enough to show that,

$$sup_{x \in S} S_1(x) = O_p(n^{-1/2}h^{-d/2} ln(n)), \tag{A.14}$$

$$sup_{x \in S} S_2(x) = O_p(h^r) + O_p(\triangle_n^{1/2}), \tag{A.15}$$

$$sup_{x \in S} S_3(x) = O_p(n^{-1/2}h^{-d/2} ln(n)), \tag{A.16}$$

$$sup_{x \in S} S_4(x) = O_p(h^r). \tag{A.17}$$

The bias terms $S_3(x)$ and $S_4(x)$ can be treated exactly as in Härdle and Luckhaus (1984). Therefore the proofs of (A.16) and (A.17) are omitted and our proofs focus on (A.14) and (A.15). To prove

(A.15), by using the Itô Lemma to $[x^i_{n,t+1}-x^i_{n,t}][x^j_{n,t+1}-x^j_{n,t}]$, we have,

$$\begin{aligned}
S_2(x) &= Ev_n(x)-v(x) \\
&= n^{-1}h^{-d}\triangle_n^{-1}\sum_{t=1}^{n-1}E\{K(\frac{x_{n,t}-x}{h})\int_{\triangle_n}E^t[(x^j_u-x^j_{t_0+t\triangle_n})(\mu_i(x_u)-\mu_i(x_{t_0+t\triangle_n}))]du\} \\
&\quad +n^{-1}h^{-d}\triangle_n^{-1}\sum_{t=1}^{n-1}E\{K(\frac{x_{n,t}-x}{h})\int_{\triangle_n}E^t[(x^j_u-x^j_{t_0+t\triangle_n})\mu_i(x_{t_0+t\triangle_n})]du\} \\
&\quad +n^{-1}h^{-d}\triangle_n^{-1}\sum_{t=1}^{n-1}E\{K(\frac{x_{n,t}-x}{h})\int_{\triangle_n}E^t[(a_{ij}(x_u)-a_{ij}(x_{t_0+t\triangle_n}))]du\} \\
&\quad +n^{-1}h^{-d}\sum_{t=1}^{n-1}E\{K(\frac{x_{n,t}-x}{h})a_{ij}(x_{t_0+t\triangle_n})\}-a_{ij}(x)f(x) \\
&\equiv S_{21}(x)+S_{22}(x)+S_{23}(x)+S_{24}(x). \qquad \text{(A.18)}
\end{aligned}$$

Using the same method for showing $B^i_n = O(n^{-2}h^{-d}\triangle_n^{1/2})$ in Lemma 1 , $i=1,2,3,4,5$, we can prove that $S_{2i}(x)=O_p(\triangle_n^{1/2})$ for $i=1,2,3$. We consider $S_{24}(x)$.

$$\begin{aligned}
S_{24}(x) &= \int K(u)a_{ij}(x+hu)f(x+hu)du-a_{ij}(x)f(x)-n^{-1}\int K(u)a_{ij}(x+hu)f(x+hu)du \\
&= O_p(h^r), \qquad \text{(A.19)}
\end{aligned}$$

where we use the Taylor expansion of the function $a_{ij}(x+hu)f(x+hu)$ at x. Therefore, (A.15) holds from $S_{2i}(x)=O_p(\triangle_n^{1/2})$, for $i=1,2,3$, and (A.19). To prove (A.14), let $Y_{n,t}\equiv\triangle_n^{-1}[x^i_{n,t+1}-x^i_{n,t}][x^j_{n,t+1}-x^j_{n,t}]$ and $v_n(x)\equiv v_n^+(x)+v_n^-(x)$, where $v_n^+(x)\equiv(nh^d)^{-1}\sum_t K(\frac{x^i_{n,t+1}-x}{h})Y_{n,t}I[|Y_{n,t}|\geq ln(n)]$. Let $z_{n,t}\equiv exp(\alpha Y_{n,t})$, where $\alpha>0$. For the Jensen inequality, Schwarz inequality, and Lemma 1, we have $E(z_{n,t})\leq e^{E(\alpha Y_{n,t})}\leq e^{\triangle_n^{-1}[E(x^i_{n,t+1}-x^i_{n,t})^2]^{1/2}[E(x^j_{n,t+1}-x^j_{n,t})^2]^{1/2}}\leq C$. Then, (A.14) is shown by using Theorem 3.3.2 (Remark 3.3.4) in Györfi et al. (1989).

(A.12) can be shown by following the similar arguments as in proving (A.11). The detailed proof of (A.12) is not incorporated here, but it is available from the author upon request.

Lemma 3. Under Assumptions 1-6 and the null hypothesis, I_{nij} can be written as

$$I_{nij}=\bar{I}_{nij}+o_p((nh^{d/2})^{-1}), \qquad \text{(A.20)}$$

where $\bar{I}_{nij} = \int [(\hat{a}_{ij}(x) - a_{ij}(x))\hat{f}(x)]^2 w(x)dF(x)$.

Proof of Lemma 3: The detailed proof of Lemma 3 is omitted because Lemma 3 can be proven by following a similar method to the one for proving Lemma 3 in Li (2007).

Proof of part (i) of Theorem 1: we will complete the proof of (i) of Theorem 1 by showing: (a) $\bar{J}_{nij} \equiv nh^{d/2}[\bar{I}_{nij} - r_{nij}(d)] \to N[0, v^2]$ in distribution, and (b) $v_{nij}^2 \to v^2$ in probability.

Proof of (a): The CLT in Fan and Li (1999) for degenerate U-statistics is extended to the triangular arrays of random variables $\{X_{n,t}, t \leq n\}$ in Li (2007); that is, if the Assumptions (A1)-(A3) in Fan and Li (1999) are satisfied, then $\sqrt{2}U_n/(n\sigma_n) \to N(0,1)$ is in distribution as $n \to 0$. Note that the Assumptions (A1)-(A3) in Fan and Li (1999) are said to be satisfied by $\{X_{n,t}\}$ if the conditions in Assumptions (A1)-(A3) in Fan and Li (1999) as calculated by every row of $\{X_{n,t}\}$ are satisfied.

Let $\bar{s}_n(z_{n,l}, x, \theta_0) = s_n(z_{n,l}, x, \theta_0) - Es_n(z_{n,l}, x, \theta_0)$, where $z_{n,l} = (x_{n,l}^i, x_{n,l+1}^i, x_{n,l}^j, x_{n,l+1}^j,)$. We decompose $\bar{I}_{nij}$ according to,

$$\begin{aligned}
\bar{I}_{nij} &= 2\sum_{1\leq l<l'\leq n}\int \bar{s}_n(z_{n,l}, x, \theta_0)\bar{s}_n(z_{n,l'}, x, \theta_0)w(x)dF(x) \\
&\quad + \sum_{l=1}^{n}\int s_n^2(z_{n,l}, x, \theta_0)w(x)dF(x) \\
&\quad + 2(n-1)\sum_{l=1}^{n}\int \bar{s}_n(z_{n,l}, x, \theta_0)E(s_n(z_{n,l}, x, \theta_0))w(x)dF(x) \\
&\quad + n(n-1)\int [Es_n(z_{n,l}, x, \theta_0)]^2 w(x)dF(x) \\
&\equiv \bar{I}_{nij}^{11} + \bar{I}_{nij}^{12} + \bar{I}_{nij}^{13} + \bar{I}_{nij}^{14}. \qquad \text{(A.21)}
\end{aligned}$$

We will show under our assumptions that $\bar{I}_{nij}^{11}$ is asymptotically normal in distribution, and $\bar{I}_{nij}^{13}$ and $\bar{I}_{nij}^{14}$ are asymptotically negligible in probability, while $\bar{I}_{nij}^{12}$ gives a bias term. First, we prove that $nh^{d/2}(\bar{I}_{nij}^{12} - E\bar{I}_{nij}^{12}) \to 0$ in probability under the null hypothesis.

$$var(\bar{I}_{nij}^{12}) = \sum_{l=1}^{n} var(\int s_n^2(z_{n,l}, x, \theta_0)w(x)dF(x))$$

$$
\begin{aligned}
&+2\sum_{1\le l<l'\le n}\{E[\int s_n^2(z_{n,l},x,\theta_0)w(x)dF(x)\int s_n^2(z_{n,l'},x,\theta_0)w(x)dF(x)] \\
&-\int Es_n^2(z_{n,l},x,\theta_0)w(x)dF(x)\int Es_n^2(z_{n,l'},x,\theta_0)w(x)dF(x)\} \\
&= \bar{I}_{nij}^{12}(1)+\bar{I}_{nij}^{12}(2). \qquad (\Lambda.22)
\end{aligned}
$$

Changing the variable in $(x_{n,l}-x)/h=u$ and using $(\Lambda.1)$, we have,

$$
\begin{aligned}
n^{-1}\bar{I}_{nij}^{12}(1) &\le E[\int s_n^2(z_{n,l},x,\theta_0)w(x)dF(x)]^2 \\
&= E\{\int (nh^d\triangle_n)^{-2}K_l^2(x)[(x_{n,l+1}^j-x_{n,l}^j)^2(x_{n,l+1}^i-x_{n,l}^i)^2 \\
&- 2a_{ij}(x)\triangle_n(x_{n,l+1}^j-x_{n,l}^j)(x_{n,l+1}^i-x_{n,l}^i)+a_{ij}^2(x)\triangle_n^2]w(x)dF(x)\}^2 \\
&\le C((n\triangle_n)^{-4}h^{-2d})E\{[\int K^2(u)w(x_{n,l}+hu)f(x_{n,l}+hu)du]^2(x_{n,l+1}^j-x_{n,l}^j)^4(x_{n,l+1}^i-x_{n,l}^i)^4\} \\
&+ C((n^2h)^{-2}\triangle_n^{-2})E\{[\int K^2(u)a_{ij}(x_{n,l}+hu)w(x_{n,l}+hu,\theta)f(x_{n,l}+hu)du]^2 \\
&\quad\times(x_{n,l+1}^j-x_{n,l}^j)^2(x_{n,l+1}^i-x_{n,l}^i)^2\} \\
&+ C(n^2h^d)^{-2}E\{\int K^2(u)a_{ij}^2(x_{n,l}+hu)w(x_{n,l}+hu)f(x_{n,l}+hu)du\}^2 \\
&= O(n^4h^{2d})^{-1}. \qquad (\Lambda.23)
\end{aligned}
$$

Under Assumption 3, the observed data sequence $\{x_{n,l}\}$ satisfies the absolutely regular condition with mixing coefficient $\beta_{n,\tau}=O(\lambda^\tau)$, where $0<\lambda^\tau<1$. Let $\tau=[blog(n)]$ and $\kappa=log\lambda$, where b is a sufficiently large positive constant. Then we have $\beta_{n,\tau}=O(\lambda^\tau)=O(\lambda^{-bklog_\lambda n})=O(n^{-b\kappa})$.

For $\bar{I}_{nij}^{12}(2)$, we consider two different cases: (a) $\{|l-l'|>\tau+1\}$ and (b) $\{|l-l'|\le\tau+1\}$. We use EB_a and EB_b to denote cases (a) and (b), respectively. By Schwarz's Inequality and Lemma 1 in Li (2007), we have,

$$
\begin{aligned}
EB_a+EB_b &\le Ch^{-d}\beta_{n,\tau}^{1/2}+C\tau n^{-3}h^{-2d} \\
&= O(h^{-d}n^{-b\kappa/2})+O(\tau n^{-3}h^{-2d}). \qquad (\Lambda.24)
\end{aligned}
$$

From $(\Lambda.22)$-$(\Lambda.24)$ and Chebyshev's inequality, it follows that $nh^{d/2}(\bar{I}_{nij}^{12}-E\bar{I}_{nij}^{12})=o_p(1)$. From

$(A.3), (A.4), (A.10) - (A.12)$, and CLT, we know that $nh^{d/2}[\bar{I}_{nij}^{12} - r_{nij}(d)] = o_p(1)$, which indicates that $nh^{d/2}[\bar{I}_{nij} - r_{nij}(d)] = (nh^{d/2})[\bar{I}_{nij}^{11} + \bar{I}_{nij}^{13} + \bar{I}_{nij}^{14}] + o_p(1)$. To prove that $nh^{d/2}\bar{I}_{nij}^{13} = o_p(1)$, we first evaluate

$$
\begin{aligned}
& E[\int \bar{s}_n(z_{n,l}, x, \theta_0) E(s_n(z_{n,l}, x, \theta_0)) w(x) dF(x)]^2 \qquad (A.25) \\
= & \int\int E[s_n(z_{n,l}, x, \theta_0) s_n(z_{n,l}, y, \theta_0)] E[s_n(z_{n,l}, x, \theta_0)] \\
& \times E[s_n(z_{n,l}, y, \theta_0)] w(x) w(y) dF(x) dF(y) \\
& - \{\int [E s_n(z_{n,l}, x, \theta_0)]^2 w(x) dF(x)\}^2 \\
= & O(n^{-4} h^{-d} \triangle_n + n^{-4} h^{2r-d}), \qquad (A.26)
\end{aligned}
$$

which is derived from Schwarz's inequality, $(A.2), (A.3)$, and $(A.4)$. Because of $E\bar{s}_n(z_{n,l}, x, \theta_0) = 0$, it follows that $nh^{d/2}\bar{I}_{nij}^{13} = O_p(1)$ by $(A.26)$ and Chebyshev's Inequality.

For $\bar{I}_{nij}^{14}$, by (A.2) we have,

$$
\begin{aligned}
nh^{d/2}\bar{I}_{n1}^{14} &= n^2(n-1)h^{d/2} \int [E s_n(z_{n,l}, x, \theta_0)]^2 w(x) dF(x) \\
&= n^2(n-1)h^{d/2}(O(n^{-2}\triangle_n) + O(n^{-2}h^{2r})) \\
&= O(nh^{d/2}\triangle_n) + O(nh^{d/2+2r}) = o(1).
\end{aligned}
$$

According to Lemma 4 in Li (2007), to prove that $\bar{I}_{nij}^{11}$ is asymptotically normal in distribution, we will only prove that the Assumptions (A1)-(A3) in Fan and Li (1999) are satisfied under the conditions in Theorem 1.

Let $H(z_{n,j}, z_{n,k}) \equiv \int \bar{s}_n(z_{n,j}, x, \theta_0) \bar{s}_n(z_{n,k}, x, \theta_0) w(x) dF(x)$ and $\{\tilde{z}_{n,j}\}_{j=1}^n$ be an i.i.d. sequence having the same marginal distribution as $\{z_{n,j}\}_{j=1}^n$. Then we have,

$$
\begin{aligned}
\sigma_n^2 &\equiv E[H^2(\tilde{z}_{n,1}, x, \theta_0, \tilde{z}_{n,2}, x, \theta_0)] \\
&= \int\int E[s_n(\tilde{z}_{n,1}, x, \theta_0) r_n(\tilde{z}_{n,1}, y, \theta_0)] E[s_n(\tilde{z}_{n,2}, x, \theta_0) s_n(\tilde{z}_{n,2}, y, \theta_0)]
\end{aligned}
$$

$$
\begin{aligned}
&\times w(x)w(y)dF(x)dF(y) \\
&-2\int\int E[s_n(\tilde{z}_{n,1},x,\theta_0)s_n(\tilde{z}_{n,1},y,\theta_0)]E[s_n(\tilde{z}_{n,2},x,\theta_0)]E[s_n(\tilde{z}_{n,2},y,\theta_0)] \\
&\times w(x)w(y)dF(x)dF(y) \\
&+\int\int E[s_n(\tilde{z}_{n,1},x,\theta_0)]E[s_n(\tilde{z}_{n,1},y,\theta_0)]E[s_n(\tilde{z}_{n,2},x,\theta_0)]E[s_n(\tilde{z}_{n,2},y,\theta_0)] \\
&\times w(x)w(y)dF(x)dF(y) \\
= \;& (n^4h^d)^{-1}\int [a_{ii}(x)a_{jj}(x)+a_{ij}^2(x)]^2 f^4(x)w^2(x)dx \\
&\times \int [\int K(u)K(w+u)du]^2 dw + o((n^4h^d)^{-1}). \qquad (A.27)
\end{aligned}
$$

The detailed proof of the following results is not incorporated to conserve space, but it is available upon request.

$$
\begin{aligned}
\mu_{n4} \equiv \;& \int\int\int\int E[\bar{s}_n(z_{n,1},x,\theta_0)\bar{s}_n(z_{n,1},\bar{x},\theta_0)\bar{s}_n(z_{n,1},y,\theta_0)\bar{s}_n(z_{n,1},\bar{y},\theta_0)] \\
&\times E[\bar{s}_n(z_{n,2},x,\theta_0)\bar{s}_n(z_{n,2},\bar{x},\theta_0)\bar{s}_n(z_{n,2},y,\theta_0)\bar{s}_n(z_{n,2},\bar{y},\theta_0)] \\
&\times w(x)w(\bar{x})w(y)w(\bar{y})dxd\bar{x}dyd\bar{y} \\
&= O((n^8h^{3d})^{-1}).
\end{aligned}
$$

$$
\gamma_{n11} \equiv max_{t\neq s,t'\neq s'} E[H(z_{n,t},z_{n,s})H(z_{n,t'},z_{n,s'})] = O(n^{-4}h^{4d/\eta-4d}), \qquad (A.28)
$$

where $1<\eta=(1-\xi^{-1})^{-1}<4/3$, and ξ is slightly larger than 2.

$$
\gamma_{n22} \equiv max_{t\neq s,t'\neq s'} E[H^2(z_{n,t},z_{n,s})H^2(z_{n,t'},z_{n,s'})] = O(n^{-8}h^{4d/\eta-6d}), \qquad (A.29)
$$

$$
\gamma_{n13} \equiv max_{t\neq s,t'\neq s'} E[H(z_{n,t},z_{n,s})H^3(z_{n,t'},z_{n,s'})] = O(n^{-8}h^{4d/\eta-6d}), \qquad (A.30)
$$

where $1<\eta<4/3$.

$$
\tilde{\gamma}_{n22} \equiv max_{t\neq s,t'\neq s'} E[H^2(\tilde{z}_{n,1},\tilde{z}_{n,2})H^2(\tilde{z}_{n,1},\tilde{z}_{n,3})] = O(n^{-8}h^{2d/\eta-4d}), \qquad (A.31)
$$

where $1 < \eta < 2$.

$$\tilde{\gamma}_{n14} \equiv max_{t \neq s} \int \{E[H(z, Z_{n,t}) H(z, Z_{n,s})]\}^2 dF(z) = O(n^{-8} h^{5d/\eta - 5d}), \qquad \text{(A.32)}$$

where $1 < \eta < 5/3$. Summarizing the above results, we have $\sigma_n^2 = O(n^{-4} h^{-d})$, $\mu_{n4} = O((n^8 h^{3d})^{-1})$, $\gamma_n \equiv max\{\gamma_{n11}, \tilde{\gamma}_{n22}, \tilde{\gamma}_{n14}\} = O(n^{-4} h^{4d/\eta - 4d})$, and $v_n = max\{\gamma_{n22}, \gamma_{n13}\} = O(n^{-8} h^{4d/\eta - 6})$. These results imply (A1)(i)-(iii) in Fan and Li (1999).

Define $G_n(x, y) \equiv E[H(Z_1, z_t) H(Z_1, z_s)]$. Then we have,

$$\sigma_G^2 \equiv E[G_n^2(Z_{n,t}, Z_{n,t})] = O(n^{-8} h^{3d/\eta - 4d}),$$

$$\mu_{nG2} \equiv max_{t \neq s} \int G^2(Z_{n,t}, Z_{n,s}) dQ^n(Z_{n,t}, Z_{n,s}) = O(n^{-8} h^{3d/\eta - 4d}),$$

where $1 < \eta < 3/2$.

$$\begin{aligned} \gamma_{nG11} \equiv\ & max\{max_{s \neq s' \neq s''} |E[G(z_{n,s}, z_{n,s}) G(z_{n,s'}, z_{n,s''})]|, \\ & max_{s \neq s' \neq s''} |E[G(z_{n,s}, z_{n,s'}) G(z_{n,s}, z_{n,s''})]|, \\ & max_{s \neq s' \neq s'' \neq s'''} |E[G(z_{n,s}, z_{n,s'}) G(z_{n,s''}, z_{n,s'''})]|\} = O(n^{-8} h^{3d/\eta - 4d}), \end{aligned}$$

where $1 < \eta < 3/2$. Thus, (A2)(i)-(iii) in Fan and Li (1999) are satisfied.

Finally, it is easy to show that M_n in Fan and Li (1999) is bounded by a positive constant. Also $\sigma_n^2 = O(\frac{1}{n^4 h^d})$ and $\beta_{n,\tau} = O(n^{-b\kappa})$. Hence, we have $\tau^2 n^2 \beta_{n,\tau}^{1/2} / \sigma_n^4 = o(1)$ provided we make b sufficiently large, which implies $n^2 \beta_{n,\tau}^{1/2} = o(1)$. Thus, (A3)(i) and (ii) in Fan and Li (1999) are all satisfied. Hence, we get $\bar{I}_{nij}^{11} / (\sqrt{2} n \sigma_n) \to N(0,1)$ in distribution, which indicates that $\bar{J}_{nij} \to N[0, v^2]$ in distribution.

Proof of (b): from Assumption 6, Lemma 2, and its proof, we have,

$$inf_{x \in S} f(x) > 0, \qquad \text{(A.33)}$$

$$sup_{x \in S} |[\hat{a}_{ii}(x_{n,t}) \hat{f}(x_{n,t}) \hat{a}_{jj}(x_{n,t} \hat{f}(x_{n,t})]^2 - [a_{ii}(x_{n,t}) f(x_{n,t}) a_{jj}(x) f(x)]^2| = o_p(1). \qquad \text{(A.34)}$$

From (A.33), (A.34), and Lemma 2, it follows that,

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\Big|\frac{[\hat{a}_{ii}(x_{n,t})\hat{f}(x_{n,t})\hat{a}_{jj}(x_{n,t})\hat{f}(x_{n,t})+\hat{a}_{ij}^2(x_{n,t})\hat{f}^2(x_{n,t})]^2 w^2(x_{n,t})}{\hat{f}(x_{n,t})} \\
&\quad -\frac{[a_{ii}(x_{n,t})f(x_{n,t})a_{jj}(x_{n,t})f(x_{n,t})+a_{ij}^2(x_{n,t})f^2(x_{n,t})]^2 w^2(x_{n,t})}{f(x_{n,t})}\Big| \\
\leq\ & sup_{x\in S}\Big|\frac{[\hat{a}_{ii}(x)\hat{f}(x)\hat{a}_{jj}(x)\hat{f}(x)+\hat{a}_{ij}^2(x)\hat{f}^2(x)]^2-[a_{ii}(x)f(x)a_{jj}(x)f(x)+a_{ij}^2(x)f^2(x)]^2}{f(x)}w^2(x)\Big| \\
&\quad +sup_{x\in S}\Big|\frac{[\hat{a}_{ii}(x)\hat{a}_{jj}(x)+\hat{a}_{ij}^2(x)]^2\hat{f}^4(x)(\hat{f}(x)-f(x))}{f(x)\hat{f}(x)}w^2(x)\Big| \\
=\ & o_p(1),
\end{aligned}$$

which indicates

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\frac{[\hat{a}_{ii}(x_{n,t})\hat{a}_{jj}(x_{n,t})+\hat{a}_{ij}^2(x_{n,t})]^2\hat{f}^4(x_{n,t})w^2(x_{n,t})}{\hat{f}(x_{n,t})} \\
=\ & \frac{1}{n}\sum_{t=1}^{n}\frac{[a_{ii}(x_{n,t})a_{jj}(x_{n,t})+a_{ij}^2(x_{n,t})]^2 f^4(x_{n,t})w^2(x_{n,t})}{f(x_{n,t})}+o_p(1) \\
=\ & \int[a_{ii}(x)a_{jj}(x)+a_{ij}^2(x)]^2 f^4(x)w^2(x)dx+o_p(1),
\end{aligned}$$

by the law of large numbers. Therefore, we have $v_{nij}^2 = v^2 + o_p(1)$.

Proof of part (ii) of Theorem 1: given assumptions 1-6, we can show, by using similar arguments to those for the proof of part (a) in (i) of Theorem 1 that, under H_a^{ij}, $I_{nij} = \int[(a_{ij}(x) - a_{ij}^0(x,\theta^*))f(x)]^2 w(x)dF(x) + o_p(1)$, where $\theta^* \in \Theta$, and $(\theta_n - \theta^*) = O_p(\frac{1}{\sqrt{n}})$. Hence, we have $J_{nij} = O_p(nh^{d/2})$.